Project Horizon — a wake-up call

Research into the effects of sleepiness on the cognitive performance of maritime watchkeepers under different watch patterns, using ships' bridge, engine and liquid cargo handling simulators.

Research report 2012

with the cooperation of

SEVENTH FRAMEWORK PROGRAMME

HORIZON

www.project-horizon.eu

Part EU-funded Project No. FP7 234000

Contents

List of abbreviations in the text

ANOVA	Analysis of Variance
COLREGS	International Regulations for the Prevention of Collisions at Sea
CPA	Closest Point of Approach
EEG	Electroencephalogram
ECG	Electrocardiogram
EOG	Electrooculogram
HRV	Hearth rate variability
KSS	Karolinska Sleepiness Scale
KDT	Karolinska Drowsiness Test
LICOS	Liquid Cargo Operations Simulator
PVT	Psychomotor Vigilance Test
SD	Standard Deviation
SE	Standard Error
VLCC	Very large crude oil carrier
WMA	Warsash Maritime Academy

Project Horizon consortium

Southampton Solent University – Warsash Maritime Academy (Coordinator) (WMA)
Bureau Veritas – Marine Division, Research Dept (BV)
Chalmers Tekniska Hoegskola AB – Dept of Shipping & Marine Technology (Chalmers)
European Transport Workers' Federation – Nautilus International (ETFN)
Stockholms Universitet – Stress Research Institute (SU)
The Standard P&I Club – Charles Taylor & Co Limited (CTPI)
European Community Shipowners Associations (ECSA)
European Harbour Masters Committee (EHMC)
International Association of Independent Tanker Owners – Intertanko (INTKO)
UK Marine Accident Investigation Branch (MAIB)
UK Maritime & Coastguard Agency (MCA)

This report presents the findings of Project Horizon – a European Commission part-funded multi-partner research initiative to investigate the impact of watchkeeping patterns on the cognitive performance of seafarers. This pioneering research sought to advance understanding of seafarer fatigue through scientific analysis of data drawn from realistic working scenarios using experienced watchkeepers on ship simulators.

The report explains the reasons why the project was considered necessary and how the research was undertaken, as well as presenting the findings and research outcomes.

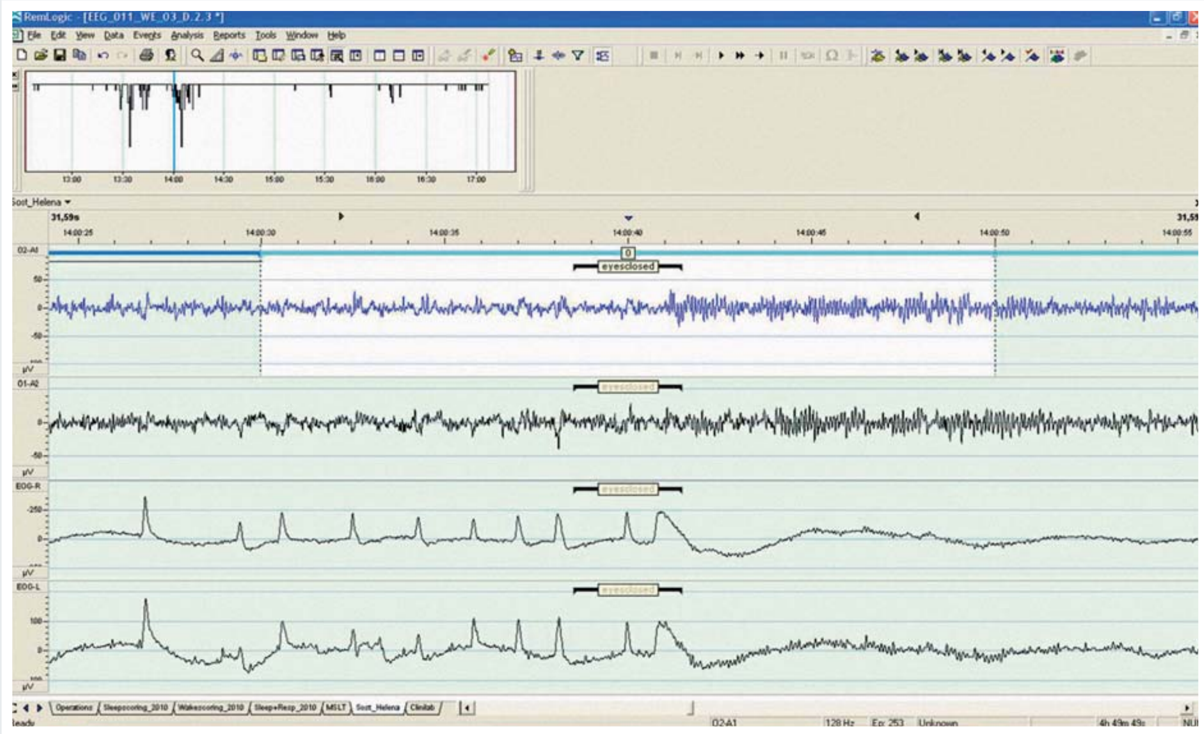
The project has taken knowledge in this area to a new level, demonstrating conclusively the links between performance degradation and certain patterns of work. The project surpasses previous subjective fatigue studies, delivering validated, scientifically and statistically robust results that can be used to help determine safer working patterns in the interests of the safety of life at sea, the safety and security of the marine transport system and the protection of the marine environment.

Introduction

Project Horizon is a major multi-partner European research study that brought together 11 academic institutions and shipping industry organisations with the agreed aim of delivering empirical data to provide a better understanding of the way in which watchkeeping patterns can affect ships' watchkeepers. The broad spread of the project partners ensured expert objectivity of the project and its results, as well as widening routes for dissemination and exploitation of the findings.

The project was established to:

- **define and undertake scientific methods for measurement of fatigue in various realistic seagoing scenarios using bridge, engineroom and cargo simulators**
- **capture empirical data on the cognitive performance of watchkeepers working within those realistic scenarios**
- **assess the impact of fatigue on decision-making performance**
- **and determine arrangements for minimising risks to ships and their cargoes, seafarers, passengers and the marine environment**



A participant's EEG brain activity measurements during a simulated watch

At the heart of the project was the extensive use of ship simulators in Sweden and the UK to examine the decision-making and cognitive performance of officers during a range of real-life, real-time scenarios of voyage, workload and interruptions. A total of 90 experienced deck and engineer officer volunteers participated in rigorous tests at Chalmers University of Technology in Göteberg, and at Warsash Maritime Academy at Southampton Solent University to measure their performance during seagoing and port-based operations on bridge, engine and liquid cargo handling simulators.

The project sought to take understanding of the issues to a new level with specialist input from some world-leading transport and stress research experts. Academic experts at WMA, Chalmers and the Stress Research Institute at Stockholm University (SU) devised the simulator runs, setting the requirements for fatigue measurement and determining performance degradation measures for watchkeepers, and SU analysed the results from the week-long programmes.

Finally, in response to the research findings, the Project Horizon partners have developed a fatigue management toolkit for the industry, which seeks to provide guidance to owners, operators, maritime regulators and seafarers to assist them in organising work patterns at sea in the safest and healthiest way possible.

Shipping is the ultimate 24/7 industry. Inherently globalised in its nature, the industry is complex, capital-intensive, increasingly technologically sophisticated and of immense economic and environmental significance. More than 80% of world trade moves by sea, almost 90% of EU external freight trade is seaborne, and some 40% of intra-EU freight is carried by shortsea shipping. Around 40% of the world fleet is beneficially controlled in the EEA and EU-registered tonnage accounts for more than 20% of the world total. An average of around four million passengers embark and disembark in 27EU ports every year – the vast majority being carried by ferries.

The increasingly intensive nature of shipping operations means that seafarers frequently work long and irregular hours. Under the International Labour Organisation regulations (social provisions) it is permissible for seafarers to work up to 91 hours a week – and, under the International Maritime Organisation's STCW 2010 amendments (safety provisions), a 98-hour working week is allowed for up to two weeks in 'exceptional' circumstances. Noise, vibration, sailing patterns, port calls, cargo handling and other activities can all reduce the ability of the seafarer to gain quality sleep during rest periods.

Fatigue is generally understood to be a state of acute mental and/or physical tiredness, in which there is a progressive decline in performance and alertness. The term is often used interchangeably with 'sleepiness', 'tiredness' and 'drowsiness'. Fatigue is often considered to be a generic term, of which sleepiness is one of the major sub-components. In this project, the emphasis has been placed upon 'sleepiness' as the most effective description of the physical and physiological conditions under examination.

Shipping is a round-the-clock industry

Seafarers are already usually covered by company, sector-specific, flag state or IMO rules banning or severely restricting alcohol use at sea. Studies have shown that around 22 hours of wakefulness will have a similar effect upon the impairment of an individual's performance as a blood-alcohol concentration of 0.10% – double the legal driving limit in most EU member states.

Laboratory research and studies in other transport modes have demonstrated that severe sleepiness (and even sleep onset) and performance deterioration is common amongst workers undertaking night shifts.

Fatigue is also an important health issue, with significant evidence to show the way in which long-term sleep loss can be a risk factor in such conditions as obesity, cardiovascular disease and diabetes.

The issue is also one of great relevance to the recruitment and retention of skilled and experienced seafarers. Reducing excessive working hours is of critical importance in delivering working conditions for maritime professionals that reflect the increasingly high levels of training and qualifications required to safely operate modern-day merchant ships.

Factors which result in fatigue include:

- **the lack of, or poor quality of, sleep**
- **working at times of low alertness**
- **prolonged work periods**
- **insufficient rest between work periods**
- **excessive workloads**
- **noise, vibration and motion**
- **medical conditions**

Project Horizon was established in response to growing concern about such issues and the increased evidence of the role of fatigue in maritime accidents. The project is therefore closely aligned to the FP7 (Sustainable Surface Transport 2008 RTD-1 call) aims of increased safety and security, and reduced fatalities.

Over the past 20 years, the shipping industry has become increasingly aware of the importance of the 'human factor' in safe shipping operations. Marine insurance statistics have shown 'human error' to be the key contributory factor in around 60% of

The Exxon Valdez disaster in 1989

The Cita grounding in 1997

The Jambo grounding in 2003

The Shen Neng 1, aground on the Great Barrier Reef in 2010

accidents – with other research suggesting that the figure is as high as 80% to 90% in the case of collisions and groundings. Such statistics have generated a growing impetus to investigate the nature of the human factors that may contribute to the causal chain in shipping accidents.

The increased complexity of ships' systems and the growing technological sophistication of onboard equipment have placed greater emphasis on the performance of seafarers – and watchkeepers in particular. The marked increase in the size of passenger ships and cargo vessels has also highlighted the potential for substantial loss of life or pollution in the event of an accident. Extrapolation of UK Marine Accident Investigation Branch statistics on the role of seafarer fatigue in shipping accidents between 1993 and 2003 suggests that significant economic savings could be made if the number of tiredness-related accidents is reduced.

As awareness of the importance of the human factor in shipping has grown, recognition of the role of fatigue in maritime safety has also increased. There have been a number of high-profile and often costly and damaging casualties in which seafarer fatigue has been shown as a key causal factor. These include:

- the **Exxon Valdez** tanker disaster in 1989. the US National Transportation Safety Board found that in the 24 hours prior to the grounding of the ship, the watchkeeper had only had 5 or 6 hours of sleep

- the grounding of the feeder containership **Cita** in the Isles of Scilly in March 1997, after the mate fell asleep and the ship sailed for two and a half hours with no one in control
- the grounding of the general cargoship **Jambo** in Scotland in June 2003, after the chief officer fell asleep and missed an intended change of course
- the grounding of the bulk carrier **Pasha Bulker** near the port of Newcastle in Australia in June 2007, in which an investigation report stated that 'the master became increasingly overloaded, and affected by fatigue and anxiety'
- the death of a Filipino AB in a fall onboard the Danish-flagged general cargo ship **Thor Gitta** in May 2009. Investigators who used FAID fatigue assessment software found that the seafarer's 6-on/6-off work pattern was at a score of 111 on the morning before the accident – a level considered to be in the very high range
- the grounding of the bulk carrier **Shen Neng 1** on the Great Barrier Reef in April 2010. The Australian Transport Safety Bureau investigation found that the grounding occurred because the chief mate did not alter the ship's course at the designated position. His monitoring of the ship's position was ineffective and his actions were affected by fatigue. Investigations showed that he had only two and a half hours sleep in the 38.5 hours prior to the casualty.

Concern about such incidents was also mirrored by a growing weight of evidence gathered from research among seafarers. It is generally accepted that fatigue at sea has been subjected to considerably less research than in other modes of transport or safety-critical industries, but from the 1980s onwards increasing academic attention was paid to working hours in the maritime sector – with a 1989 Medical Research Council report on hours of work, fatigue and safety at sea, by Professor ID Brown, serving as something of a watershed. In 1990, a report on shipboard crew fatigue, safety and reduced manning, by JK Pollard, ED Sussman and M Sterns noted that work at sea is characterised by longer working weeks, more non-standard work days, extensive night operations, and periods of intense effort preceded by periods of relative inactivity.

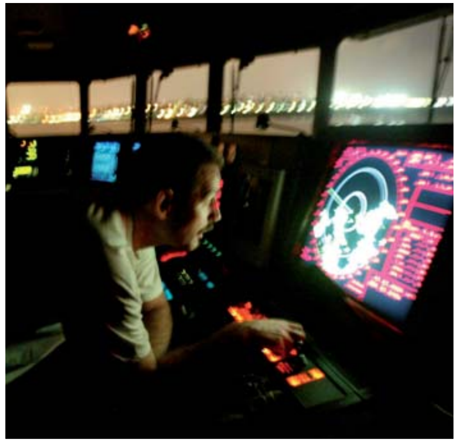
In 1995, the UK National Union of Marine Aviation & Shipping Transport Officers (NUMAST) published the result of a survey of 1,000 officers. Just over three-quarters of those surveyed said they believed that fatigue had increased significantly in the previous three to 10 years. In a further survey of 563 members, NUMAST found 50% reporting that they worked more than 85 hours a week.

A 2006 report on one of the most extensive research projects, carried out by the Centre for Occupational and Health Psychology at Cardiff University, found evidence that as many as one in four watchkeepers reported having fallen asleep on watch. As many as 53% of respondents reported having no opportunity to have six hours of uninterrupted sleep. A Swedish survey carried out in 2008 and 2010 showed that about 70% of officers had nodded off on watch one or more times during their career.

Another significant research study was published by the UK Marine Accident Investigation Branch (MAIB) in 2004. This analysed the role of fatigue in 66 collisions, near-collisions, groundings and contacts investigated between 1989 and 1999. Fatigue was considered to be a contributory factor to 82% of the groundings in the study which occurred between 0000 and 0600 and was also a major causal factor in the majority of collisions.

This latter point was also highlighted in research published by the Karolinska Institute in Sweden in 2004, which found levels of sleepiness to be highest during the 00:00-06:00hrs watch period.

In 2005, a report published by TNO in the Netherlands, recommended the setting up of a framework for the development of a fatigue management programme or tool to help shipping companies to take measures to manage fatigue.



Shipping is a round-the-clock industry

Other seafarer fatigue studies have also highlighted such factors as:

- **the long working hours experienced by many crew members**
- **problems in gaining quality sleep**
- **the impact of watchkeeping patterns: notably six hours-on/six hours-off**
- **stress and workloads**
- **frequent port calls and associated cargo work**
- **tour lengths**

Against this background, Project Horizon seeks to address the marked concerns over the increasing human, financial and environmental impact of maritime accidents which frequently cite fatigue as a contributory cause. This is an issue of critical importance at a time when the high demand for shipping capacity has led to national and international shortages of well-qualified and experienced seafarers.

Project Horizon research has been based on very rigorous scientific principles, involving unprecedented and cutting-edge use of deck, engine and cargo handling simulators to create realistic seven-day simulated voyage scenarios for the volunteer officers.

The study was focussed upon two of the most common watch schedules used at sea: six hours on watch followed by six hours off (6-on/6-off) and four hours on followed by eight hours off (4-on/8-off). The 6-on/6-off pattern is most common on smaller ships, often operating in shortsea and coastal trades and often operating with just two officers onboard.

The simulator voyage plans were designed to ensure a high degree of authenticity, including variable workloads, port visits, mandatory reporting points, and passing traffic.

The studies were carried out using the simulators at Warsash Maritime Academy in the UK and Chalmers Technical University in Sweden. At Warsash, the effects of the 6-on/6-off schedule were observed for deck and engine watchkeepers, whilst at Chalmers the tests examined the effects of 4-on/8-off and 6-on/6-off watches on deck watchkeepers only.

Before the simulator runs began at Chalmers and Warsash, extensive pilot tests were conducted to ensure the methodology was right and a Simulation Protocol Handbook was produced.

A total of 90 officers were recruited to undertake the simulated voyages. All those taking part were appropriately qualified and experienced deck and engineer officers from west and east Europe, Africa and Asia. The mix of nationalities and gender (87 males and three women) provided a representative cross-section from the industry and all participants were required to be in good health, with no sleep disorders. The volunteers were recruited through advertisements and crewing agencies as if they were going to sea and during the tests they lived as close to a shipboard life as possible – in institutional-style cabin accommodation at WMA and onboard an accommodation vessel at Chalmers. During the runs, there were a number of imposed restrictions and participants were allowed up to four cups of coffee a day, and no alcohol was permitted.

Background characteristics of the Project Horizon participants

	Chalmers		Warsash	
	4-on/8-off	**6-on/6-off**	**Deck**	**Engineroom**
n	30	19	20	20
# of men	29	18	20	19
# of women	1	1	0	1
Age (years)	30 ±6	34 ±12	30 ±7	32 ±8
Married	60%	47%	65%	65%
Children	37%	42%	15%	45%
Years at sea	7	11	8	9
Weight (kg)	82 ±9	87 ±11	79 ±8	75 ±9
Height (m)	1.81 ±0.06	1.82 ±0.06	1.75 ±0.05	1.73 ±0.08

Table 1

The total time spent 'working' during the week-long simulator runs was 64 hours for those on 4-on/8-off and 90 hours for 6-on/6-off participants (including at Chalmers an interrupted off-watch period). In that experiment, participants were randomly assigned to a watch system and a simulator and were told in advance that one of their free watches would be interrupted – although they were not told which one it would be. During the interrupted off-watch period, participants were supervised and had to undertake a mix of cargo operations simulator work and 'paperwork', including reading and watching the TV. They were not allowed to sleep during this period. This element of the programme was introduced to simulate real-world conditions, in which work patterns may be interrupted by such factors as port visits, inspections, cargo work, drills and emergencies. To balance the experiment design, one watch system had this disturbed off-watch period in the first part of the week, and the second session with the same watch system had it in the second part of the week.

The test methodology was rigorous. Cameras tracked and recorded participants' every movement on watch, producing an enormous database of activity, while supervisors were able to observe remotely on CCTV monitors.

Instructors were able to oversee the 'voyages', not only monitoring performance but also acting as 'masters' and 'chiefs' during handovers and in cases where intervention has been required to prevent an accident. The policy was one of minimal intervention, but instructors could not allow a collision, grounding or other major incident to occur as this would have prevented the completion of the exercise under experimentally controlled conditions.

The following data were collected:

- **Actigraphy** – participants wore the Actiwatch, a device that measures acceleration and enables physical activity and sleep duration to be calculated
- **Electroencephalogam (EEG)**, electrooculogram (EOG), and electrocardiogram (ECG) – recordings of brain activity, eye movements and heart rates
- **Psychomotor Vigilance Test (PVT)** performed, using standard hand-held equipment, before and after each watch. The test involved participants having to press a button to record when they see a target presented on a screen at random intervals. Each test lasted approximately five minutes and the reaction time, the number of lapses, and the mean reaction time were all recorded and stored on the device
- **Karolinska Sleepiness Scale (KSS)**
- **Karolinska Drowsiness Test (KDT)** – administers at the end of a watch, when participants' EEG measurements were taken as they were asked to stare at a black spot on a wall for five minutes and then to close their eyes for five minutes
- **Stress scores**

A 'near-miss' incident during the simulated voyage

A participant undertakes the Karolinska drowsiness test

- **Stroop test** – in which participants were sat at a laptop computer on which the names of two different colours (green and red) were shown on the screen. Participants had to click on the colour-name as quickly as possible, ignoring the meaning of the word displayed
- **Evaluation of general watchkeeping performance during navigation, engineroom and cargo operations**
- **Evaluation of performance in 'specific' repeatable events**
- **Demographic data (background questionnaire)**
- **Sleep and wake diary**
- **Ship's logbook**
- **Temperature in simulators and quarters**
- **Videos in all simulators**
- **Debriefing interview**

Data on participants' alertness and sleepiness was amassed using both subjective and objective research methods. The subjective information was drawn from the three diaries participants were asked to keep: a sleep diary filled in on waking up; a work diary they completed during the watch; and a wake diary completed during the off-watch period. Data collected covered:

Work diary

- **Food intake**
- **Symptoms of fatigue during work shift**
- **Work (difficult/easy)**
- **Satisfaction with own performance**
- **Workload**
- **nodding off**

Wake diary

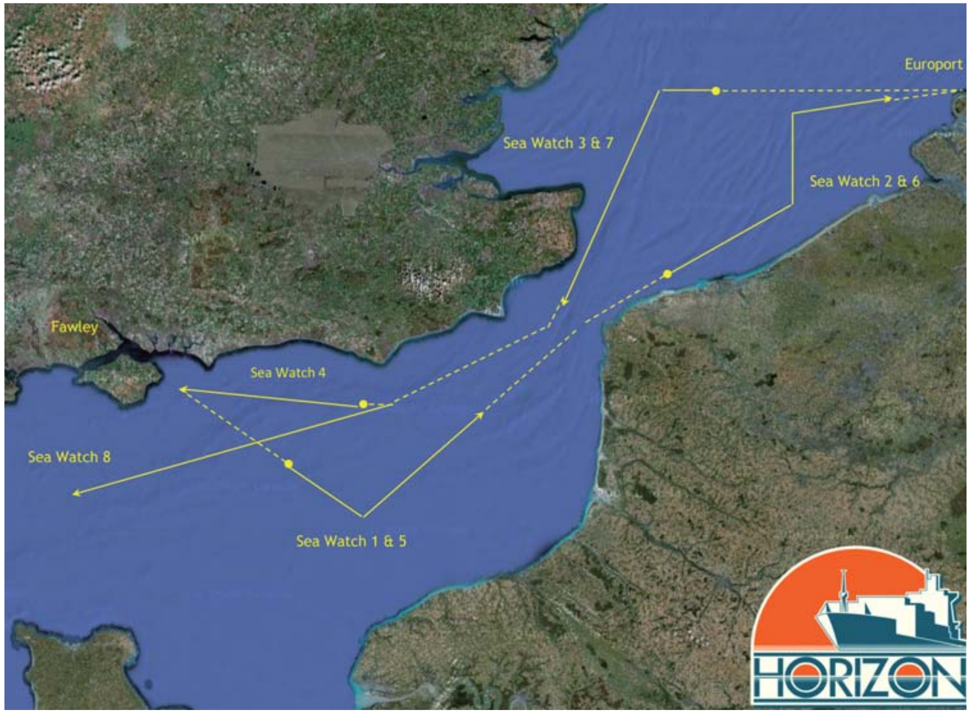
- **Food intake**
- **Type of activity during free time**
- **Symptoms of fatigue**
- **Wellbeing (health)**
- **Recuperation**

Sleep diary

- **intake of coffee**
- **intake of medications**
- **awakenings**
- **difficulty to fall asleep**
- **sleep quality**
- **waking up early**
- **easiness to get up**
- **disturbed sleep**
- **time awake during sleep period**
- **depth of sleep**
- **anxiety**
- **special occurrences**
- **reason for waking up**
- **comments**

In the watch diary, participants indicated how they felt at various points on duty using the Karolinska Sleepiness Scale. This ranges from 1 for 'extremely alert' to 9 for 'very sleepy, great effort to keep awake, fighting sleep'. This scale has been validated against road driving accidents and electroencephalogram (EEG) changes characterising sleep.

For two 24-hour periods the participants wore 10 scalp electrodes and ambulatory recorders of the EEG, which is the gold standard for measuring sleep and thus the



The simulated voyage undertaken by Project Horizon participants

absence of watchkeeper performance if it appears. They also wore Actigraph activity measuring devices to record brain and physical activity throughout the week, as well as being subjected to psychomotor vigilance tests (PVT) to check their reaction times at the beginning and end of each watch. The latter is considered the gold standard for behavioural fatigue measurement.

At two stages of the 'voyage', the participants wore 10 electrodes that measure their brain activity, over two watch periods and two sleep periods. Data obtained allows experts to analyse cognitive performance at key stages and can also show instances of 'microsleep'. Data recorded from the off-watch periods was especially valuable, as it enabled an objective picture to be obtained of exactly when participants fell asleep and the quality of the sleep they obtained.

At Chalmers, navigation simulations were carried out using two different watch schedules: 30 seafarers were assessed over 4-on/8-off schedules, and 20 were monitored on 6-on/6-off patterns. The voyage pattern was based on a simulated voyage in a small coaster and cargo simulations replicating a 210,000dwt VLCC. The data gained from these different patterns were analysed separately. The two-watch runs also included a section involving the disturbance of a single free watch, in which no sleep was allowed to enable the investigation of the effect additional workloads arising from a port visit.

At Warsash, bridge and engineroom simulators were used to investigate the effects of 6-on/6-off work patterns. Cargo handling simulations were carried out at both locations.

At Warsash, the simulators were linked up, so that the participants sailed a 17,071dwt product tanker from Fawley to Rotterdam and back again, twice, with a varied workload including cargo loading and discharge, and picking up pilots.

The simulations included some 'distinctly boring' sections as well as a number of realistic events and incidents, including:

- **keeping the ship's logbook**
- **marking positions on a chart**
- **exchanging information at the end of a watch**
- **radio communications**
- **close-quarters situations**
- **a 'man overboard' from another ship**
- **a gyro-compass error**
- **machinery alarms**

Using simulators allowed the researchers to 're-set' the voyage at the end of each watch, so that the watchkeeper coming on duty repeated the section of the voyage just completed by the previous participant. As 'handovers' were conducted by staff members acting in the role of master or chief engineer, the participants were unaware that the voyage sections were being repeated in this manner. The standard test conditions and replicated situations enabled the researchers to make valid comparisons, under statistically robust conditions, monitoring the way in which the volunteer officers reacted and how their judgement and performance were affected at different times during the week.

Volunteers' performance was also checked by a wide range of indicators – with lecturers monitoring such things as their behaviour, body language and ability to pass on 10 standard items of information at each watch handover.

During each bridge watch, participants were observed and rated by the simulator operators. The scoring system covered the general performance over the whole watch, the watch handovers, 'special' events – such as certain close-quarters situations – and 'unplanned' events – such as unintentional 'near-misses' with other vessels. The evaluation of watchkeeping performance was based on both expert rating (for example, how well the collision prevention regulations were followed) and objective scores (for example, the number and timing of positions marked on the chart).

A participant on the liquid cargo handling simulator equipment

The cargo work simulations enabled supervisors to monitor performance on a range of standard task indicators, including:

- **correct sequence of events**
- **avoidance of 'forbidden' operations**
- **control of bending moments, shear forces and list**
- **ballast handling**
- **stability control**
- **monitoring pressures and temperatures**

Similarly, engineroom performance was rated on a wide range of indicators, including:

- **standard watchkeeping duties**
- **adherence to standing orders and chief engineer's orders**
- **logbook entries**
- **communications with bridge**
- **quality of information at handovers**

- **dealing with 'incidents'** including main engine exhaust gas temperature deviation, changing over of alternators, high scavenge air temperature on main engine, boiler flame failure, high engineroom bilge level, high differential on sea water inlet strainer, fluctuating main engine fuel viscosity, and an earth fault (with request to start up other machinery while earth fault is present)

In presenting the research findings, variables that were measured once per watch were analysed using repeated measures analysis of variance (Anova) with day (1 to 7) and watch (first or second watch of the day) as 'within subject' factors and watch team (working 00:00 to 06:00 or working 06:00 to 12:00) as 'between subject' factors. Variables measured at the start and end of every watch (PVT) also included those timepoints as 'within subject' factors, and variables measured on an hourly basis (KSS and stress) included hours in watch as a 'within subject' factor.

For the Warsash runs, the analysis was carried out separately for the deck and the engineroom teams. For the Chalmers runs, analysis was carried out separately for the two watch systems (4-on/8-off and 6-on/6-off).

Karolinska Sleepiness Scale (KSS)

Sleepiness was rated every hour on the KSS self-rating scale, which has been validated against EEG measurements. The KSS scale varies from 1 to 9, with Score 1 representing highest alertness and Score being close to falling asleep. The KSS ratings are:

1. **Extremely alert**
2. **Very alert**
3. **Alert**
4. **Rather alert**
5. **Neither alert nor sleepy**
6. **Some signs of sleepiness**
7. **Sleepy, but no effort to keep alert**
8. **Sleepy, some effort to keep alert**
9. **Very sleepy, great effort to keep alert, fighting sleep**

Chalmers 4-on/8-off

Sleepiness scores were found to differ significantly between the first and the second watch of the day. The difference was highest in the team working 04:00 to 08:00 (4.1 ±0.3 versus 2.9 ±0.3) and virtually absent in the team working 08:00 to 12:00 (3.8 ±0.4 versus 3.7 ±0.3).

Sleepiness was also found to peak at the end of the watch (4.1 ±0.2), with the three-way interactions and the pattern of results indicating that maximum sleepiness is reached towards the end of the 00:00-00:04 watch, closely followed by the 04:00-08:00 watch. Lowest sleepiness scores occurred in the afternoon or early evening watches.

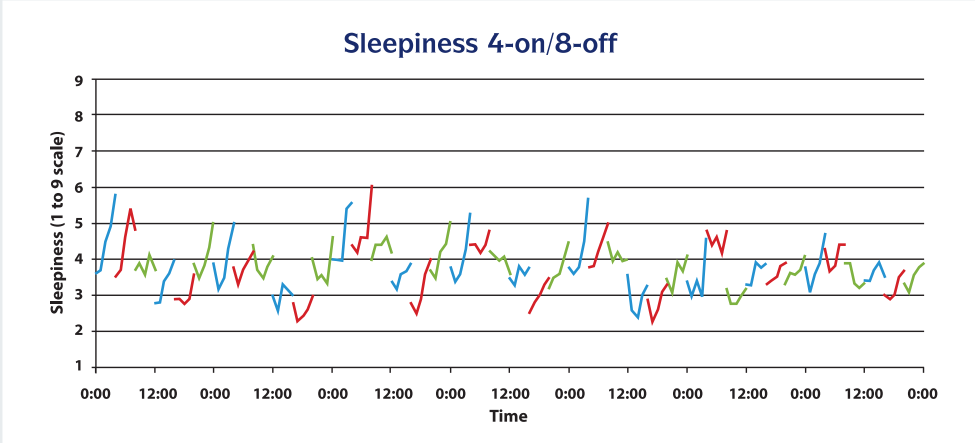


*Fig. 1 Sleepiness scores for Chalmers participants working in the 4-on/8-off bridge simulators. The team working 0000 to 0400 and 1200 to 1600 watches is indicated in **blue**; the team working 0400 to 0800 and 1600 to 2000 in **red**; and the team working 0800 to 1200 and 2000 to 0000 in **green***

Chalmers 6-on/6-off

Within subjects, sleepiness scores were found to be significantly higher during the first watch of the day than the second (4.6 ±0.2 versus 4.0 ±0.2). Sleepiness scores also differed based on the hours in watch – being lowest after one hour in watch (3.7 ±0.2) and highest at the end of the watch (5.1 ±0.3). A more complex three-way interaction between watch, hours in watch, and watch team was observed ($F(3.95, 51.40) = 10.88$, $p < 0.001$).

Effect of the off-watch disturbance

In both watch systems, the off-watch disturbance had a profound effect on sleepiness. In the 4-on/8-off system, sleepiness levels were higher during the watch following the disturbance (6.5 ±0.3) compared with the control watch (4.2 ±0.2) in the other half of the week. A similar pattern was observed in the 6-on/6-off system, with sleepiness levels being considerably higher following the off-watch disturbance (6.7 ±0.4) than during the control watch (4.6 ±0.3. No interactions were observed, indicating that the effect was similar in all watch teams. A higher rate of sleep on watch was discovered amongst participants who had experienced the disturbed off-watch period.

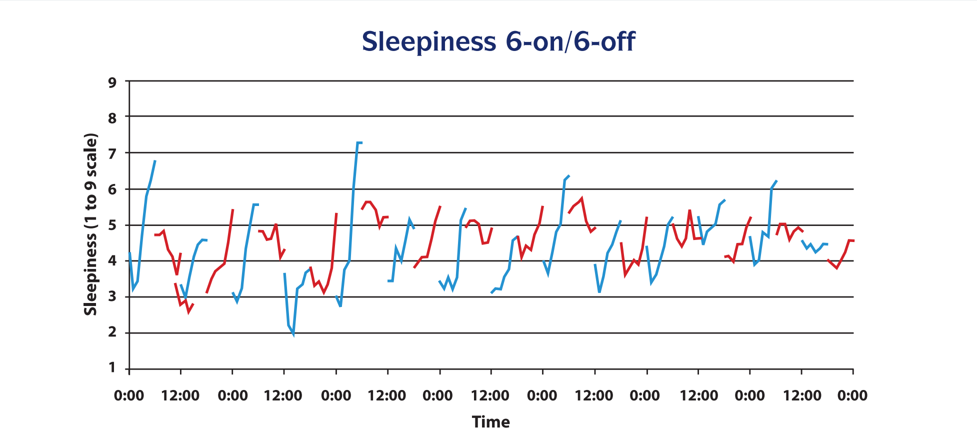


*Fig.2 Sleepiness scores for Chalmers participants working in the 6-on/6-off bridge simulators. The team working 0000 to 0600 and 1200 to 1800 is indicated in **blue**, the team working 0600 to 1200 and 1800 to 0000 in **red***

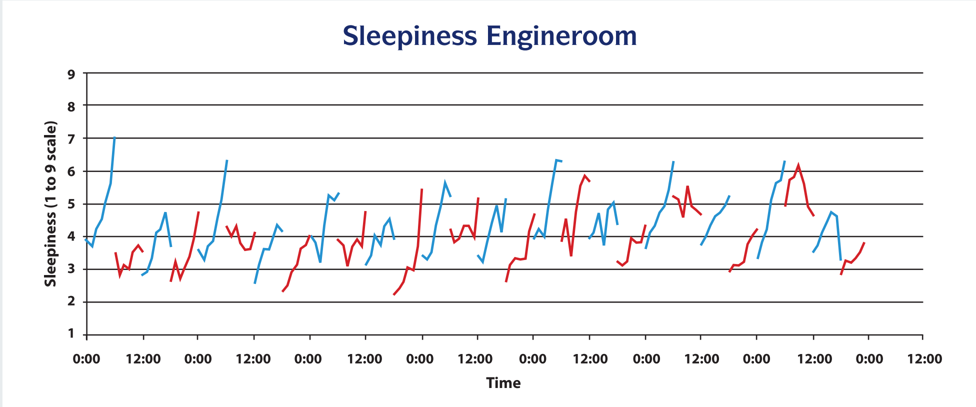


*Fig.3 Sleepiness scores for Warsash participants working in the engineroom simulators. The team working 0000 to 0600 and 1200 to 1800 is indicated in **blue**, the team working 0600 to 1200 and 1800 to 0000 in **red***

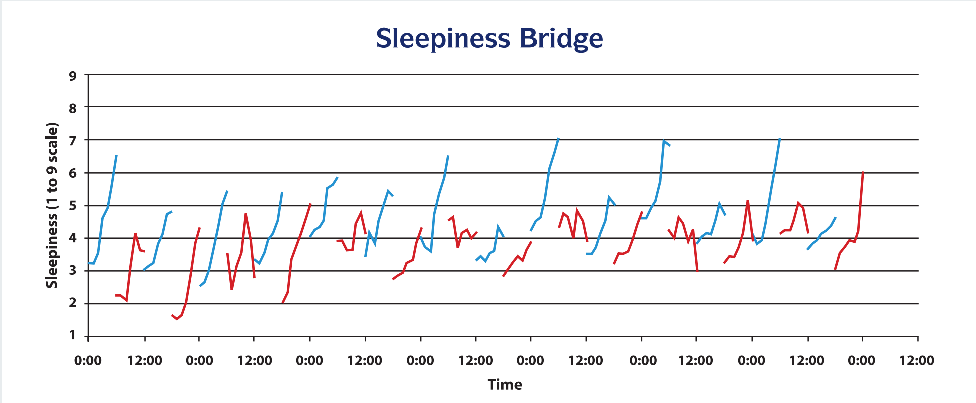


*Fig.4 Sleepiness scores for Warsash participants working in the bridge simulators. The team working 0000 to 0600 and 1200 to 1800 is indicated in **blue**, the team working 0600 to 1200 and 1800 to 0000 in **red***

4-on/8-off versus 6-on/6-off

Sleepiness levels differed between the two watch systems, being found to be higher in all watches and for all teams in the 6-on/6-off watch system (4.6 ±0.2) than in the 4-on/8-off watch system (3.9 ±0.2). Higher rates of sleep on watch were found in the 6-on/6-off teams than in the 4-on/8-off participants and sleep duration was found to be longer for those on 4-on/8-off than in the 6-on/6-off pattern.

Warsash deck

Within subjects, sleepiness scores differed significantly across the days of the week, being lowest on day 1 (3.1 ±0.3) and highest on both days 5 and 6 (4.3 ±0.4). Sleepiness scores were also found to be higher during the first watch of the day than the second (4.1 ±0.3 versus 3.6 ±0.3). Sleepiness scores also differed based on the hours in watch, ranging from 3.3 ±0.3 at the start to 4.8 ±0.3 after 5 hours in watch. Sleepiness levels were also shown to have increased during the course of the week. Daily sleep durations were found to total between 6 and 7 hours.

Warsash engineroom

Sleepiness scores differed significantly across the days of the week, being lowest on day 2 (3.5 ±0.2) and highest on day 7 (4.3 ±0.3). Scores were also found to be higher during the first watch of the day than in the second (4.2 ±0.3 versus 3.5 ±0.3). Sleepiness was found to increase during the watch period, with scores ranged from 3.3 ±0.3 at the start to 4.5 ±0.3 after 5 hours in watch. Sleepiness levels were also shown to increase during the course of the week.

Deck versus engineroom

Overall sleepiness ratings did not differ between the bridge and the engineroom.

Conclusions

- overall, more sleepiness was recorded during the first watch of the day – especially among deck teams
- sleepiness was found to increase with time in watch
- the off-watch disturbance instantly increased sleepiness
- on the whole, sleepiness levels were higher in the 6-on/6-off system than in the 4-on/8-off system
- sleepiness levels did not significantly differ between deck and engineroom
- sleepiness levels consistently peaked between 0400 and 0800
- alertness levels consistently peaked between 1400 and 1800

Stress scale

Stress was rated every hour on a 1 (very low stress – I feel very relaxed and calm) to 9 (very high stress – I feel very tense and under high pressure, on the limit to what I can manage).

Chalmers 4-on/8-off

Stress levels remained fairly low under all circumstances.

Chalmers 6-on/6-off

Stress levels remained fairly low under all circumstances.

Effect of the off-watch disturbance

In the 4-on/8-off system, stress levels were higher during the watch following the disturbance (3.7 ±0.3) than in the control watch (2.8 ±0.2) in the other half of the week. A similar effect was observed in the 6-on/6-off system, with higher stress levels following the disturbance (4.0 ±0.5) than in the control watch (2.9 ±0.2). An interaction with hours in watch was observed in the 6-on/6-off system. Following the off-watch disturbance, stress levels increased during the course of the watch, whereas such a trend was not observed during the control watch.

4-on/8-off versus 6-on/6-off

Stress levels did not differ between the two watch systems, although levels were slightly higher in the 6-on/6-off watch system (3.1 ±0.2) than in the 4-on/8-off system (2.7 ±0.2).

Warsash deck

Stress scores were found to be higher during the first watch of the day than in the second (2.8 ±0.2 versus 2.5 ±0.2). Stress scores also differed based on the hours in watch, ranging from 2.1 ±0.2 at the start of the watch to 3.1 ±0.3 after 5 hours in watch. An interaction between day and watch was observed – indicating that the effect of watch was not identical across the days of the week.

Warsash engineroom

Within subjects, stress scores differed based on the hours in watch, ranging from 3.1 ±0.3 at the start of the watch to 3.6 ±0.3 after 2 hours in watch. In addition, an interaction between watch and hours in watch was observed, indicating that the effect of hours in watch was different for the two watches.

Bridge versus engineroom

Stress ratings were higher in the engineroom than on the bridge.

A participant in the engineroom simulator at Warsash

Conclusions

- **stress levels were found to vary, but the axis along which it varied differed between the watch systems and between deck and engineroom teams**
- **overall, stress levels remained fairly low**
- **the disturbed off-watch period resulted in an immediate increase in stress levels**
- **stress levels were higher in the engineroom than on the bridge**
- **stress levels did not differ between the two watch systems**

Wake diary

Participants were asked to provide ratings on a scale of 1 (not at all) to 5 (to a great extent) on whether they had experienced any irritability, tensions, worn-out feelings, exhaustion, anxiety, or persistent fatigue, and to rate their health and whether they had got enough rest and recuperation during the last period of wakefulness.

Chalmers 4-on/8-off

No within or between subjects difference was observed in response to questions about feelings of irritability, being worn out, anxious, exhausted, or feeling persistently fatigued. However, more tensions were reported following the second watch of the day and rest and recuperation was reported to be less sufficient during this period.

Chalmers 6-on/6-off

No within or between subjects difference was observed in response to questions about feelings of irritability, being tense, worn out, anxious, exhausted, feeling persistently fatigued, or on ratings of health during the last period of wakefulness.

However, a significant difference within subjects was observed in response to the question about gaining enough rest and recuperation during the last period of wakefulness – with rest and recuperation reported to be more sufficient during the time off period following the first watch of the day than the second (2.6 ±0.2 versus 3.2 ±0.2).

Effect of the off-watch disturbance

Scores on most wake diary parameters indicated a worse state following the off-watch disturbance. For example, the category 'exhausted' obtained a rating of 2.9 ±0.3 after the disturbance vs 1.7 ±0.2 after no disturbance for 6-on/6-off. The corresponding values for 4-on/8-off were 1.8 ±0.2 vs 1.2 ±0.1. Both were highly significant, but those on the 6-on/6-off pattern were more affected – presumably because they lost a six-hour free watch, while those on 4-on/8-off 'only' lost four hours. Similar effects and ratings were seen for the category 'worn out'.

4-on/8-off versus 6-on/6-off

Two wake diary parameters differed significantly between the two watch systems. Worn out feelings were more substantial in the 6-on/6-off system (2.1 ±0.2) than in the 4-on/8-off system. The sufficiency of rest and recuperation was reported to be higher in the 4-on/8-off system (2.3 ±0.1) than in the 6-on/6-off system (3.2 ±0.3).

Warsash deck

Wake diaries showed feelings of irritability, tensions, worn out, anxiety, self-rated health and sufficiency of rest and recuperation increased during the course of the week. Tensions were higher following the second watch when rest and recuperation was stated to be less sufficient compared to the first watch.

Warsash engineroom

No significant differences within or between subjects were observed in responses to questions about irritability, tension, exhaustion, anxiety, or self-ratings of health. But feelings of persistent fatigue and insufficient rest and recuperation increased during the course of the week – being shown to be more abundant during the second watch of the day than the first.

Deck versus engineroom

Wake diary parameters did not differ between the bridge and the engineroom.

Conclusions

- wake diary outcomes indicated better time off following the first watch of the day: rest and recuperation was rated as more efficient and less negative symptoms such as tensions occurred
- outcomes got worse during the course of the week
- the disturbed free watch had adverse effects in both watch systems
- overall, more negative wake diary outcomes were reported in the 6-on/6-off system than in the 4-on/8-off system
- no differences were observed between the bridge and the engineroom

Work diary

Participants were asked whether they had experienced on a scale of 1 (not at all) to 5 (to a great extent): heavy eyelids; 'gravel eyes'; difficulties focussing; irresistible sleepiness; tired eyes; difficulties holding eyes open; impaired performance; effort to stay awake; and to rate difficulty in working; work performance; and workload; and whether they had nodded off during the watch.

Chalmers 4-on/8-off

No within or between subjects differences were observed for the parameters of gravel eyed, difficulties focussing, irresistible sleepiness, impaired performance, effort to stay awake, work performance, and nodding off.

Responses showed that the experience of heavy eyelids differed across the days and between the first and the second watch of the day, while the experience of tired eyes was reported to be higher during the first watch. Within subjects, the experience of having difficulties holding the eyes open was reported as higher during the first watch. Within subjects, self-reported work difficulty and workloads differed across the days.

Overall, the work diary parameters indicated more sleepiness and fatigue during the first watch of the day than in the second.

The ship used to provide accommodation for Chalmers participants

Chalmers 6-on/6-off

Within subjects, the experience of heavy eyelids, 'gravel eyes', difficulties focussing, irresistible sleepiness, tired eyes, difficulties holding the eyes open, was reported as higher during the first watch than the second. Self-reported nodding off was found to be higher during the first watch.

No within or between subjects difference were observed for the parameters of impaired performance, effort to stay awake, and self-rated work performance.

Within subjects, self-reported work difficulty differed across the days and was reported as being more difficult during the second watch of the day. Within subjects, self-reported workload differed across the days and a three way interaction between day, watch and watch team was observed, indicating that the effect of day was dependent on the watch of the day and that this dependency, in turn, is dependent on the watch team.

Overall, many of the work diary parameters indicated increased levels of sleepiness and fatigue during the first watch of the day compared with the second watch.

Effect of the off-watch disturbance

The effects of the off-watch disturbance compared to the control watch in the other half of the week were evident from a number of ratings. For example, the rating of 'heavy eyelids' was higher during the watch after the disturbed free watch: 2.9 ±0.2 vs 1.8 ±0.2 for 4-on/8-off and 3.3 ±0.3 vs 2.1 ±0.2 for 6-on/6-off – both highly significant. Similar results were seen for the rating 'impaired performance' (2.2 ±0.2 vs 1.5 ±0.1 for 4-on/8-off and 2.9 ±0.2 vs 1.8 ±0.1 for 6-on/6-off) as well as 'nodding off' (2.1 ±0.2 vs 1.3 ±0.1 for 4-on/8-off and 2.7 ±0.3 vs 1.6 ±0.1).

Similar variations were seen in the results for a number of other ratings of performance and fatigue. It was evident that using an off-watch period for activity rather than sleep caused major effects on perceived fatigue and performance.

Overall, the work diary parameters indicated increased sleepiness and fatigue following the free watch disturbance in both shift systems and in all watch teams.

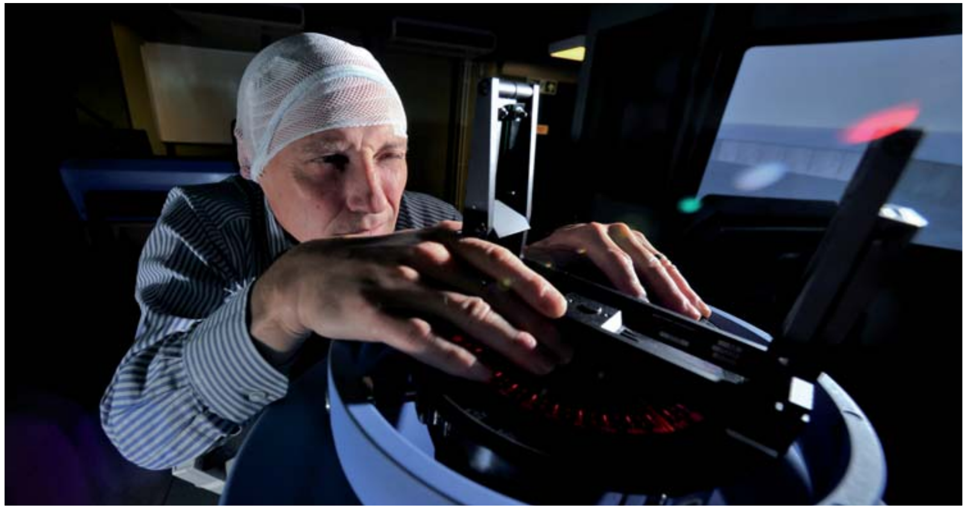
4-on/8-off versus 6-on/6-off

Several work diary parameters differed between the two watch systems.

Heavy eyelids and gravel eyes were more abundant in the 6-on/6-off system than in the 4-on/8-off system. Participants working in the 6-on/6-off system had more difficulties focussing the eyes and reported a higher incidence of tired eyes. Difficulties holding the eyes open were also more abundant in those working 6-on/6-off and those working this system also reported having to put in more effort to stay awake than those working 4-on/8-off. Subjects in the 6-on/6-off system also nodded off more frequently.

Warsash deck

Both between and within subjects, the experience of heavy eyelids was reported as higher during the first watch. Between subjects, the watch team working from 00:00 to 06:00 reported more experiences of gravelled eyes than the second watch team. Difficulties focussing the eyes and experiences of near-irresistible sleepiness, tired eyes, difficulties holding the eyes open, impaired performance, effort to stay awake, self-reported work difficulty differed across the days, and were found to be higher during the first watch. More nodding-off was reported for the first watch than during the second watch.



A participant in the bridge simulator at Warsash

Overall, most work diary parameters indicated increased levels of sleepiness and fatigue during the first watch of the day than in the second. Some parameters also indicated higher rates of sleepiness and fatigue in the team working the 00:00 to 06:00 system.

Warsash engineroom

Participants reported a higher rate of experiencing heavy eyelids, gravel eyes, difficulties focussing, irresistible sleepiness, tired eyes, difficulties holding the eyes open, impaired performance, and effort to stay awake during the first watch. Self-reported performance satisfaction levels were higher during the second watch and self-reported workload ratings were higher during the first watch. More nodding-off was reported for the first watch.

The work diary parameters indicated increased levels of sleepiness and fatigue during the first watch of the day as compared with the second. Sleepiness and fatigue symptoms also increased during the course of the week.

Deck versus engineroom

Sleepiness and fatigue-related parameters of the work diary did not differ between the bridge and the engineroom, although work difficulty and workload was found to be lower in the engineroom than on the bridge.

Overall, more sleepiness and fatigue-related symptoms were recorded during the first watch of the day. Fatigue symptoms were shown to be more abundant in the 6-on/6-off system than the 4-on/8-off system, and participants rated work difficulty and workload levels as higher on the bridge than in the engineroom.

Sleep diary

In the sleep diary, participants wrote down details including if and when they slept and how long it took for them to fall asleep (sleep latency). Questions also addressed the quality of the sleep, with subjects asked to rate their feelings at bedtime and at getting up on a scale of 1 (very alert) to 9 (very sleepy). They were also asked to state the number of cups of coffee consumed, as well as the number of sleeping pills, painkillers or any other medication that was taken.

Other questions included:

- **number of awakenings from 0 to 6 or more**
- **was it hard to fall asleep? 5 (not at all) to 1 (very hard)**
- **did you wake up too early? 5 (no) to 1 (much too early)**
- **how have you slept? 5 (very well) to 1 (very bad)**
- **was it easy to get up? 5 (very easy) to 1 (very hard)**
- **did you have disturbed sleep? 5 (not at all) to 1 (very disturbed)**
- **did you spent time awake during the period of sleep? 5 (no) to 1 (more than 1 hour)**

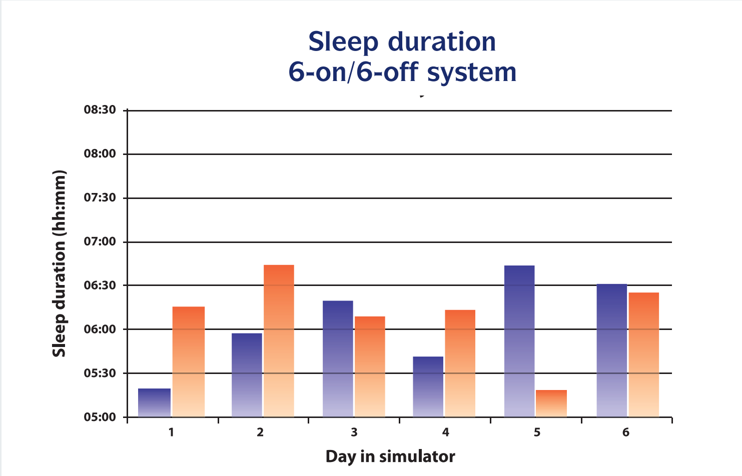


*Fig.5 Sleep duration per 24-hour period during the experimental week for Chalmers participants working the 6-on/6-off system bridge simulators. The team working 0000 to 0600 and 1200 to 1800 is indicated in **blue**, the team working 0600 to 1200 and 1800 to 0000 in **orange***

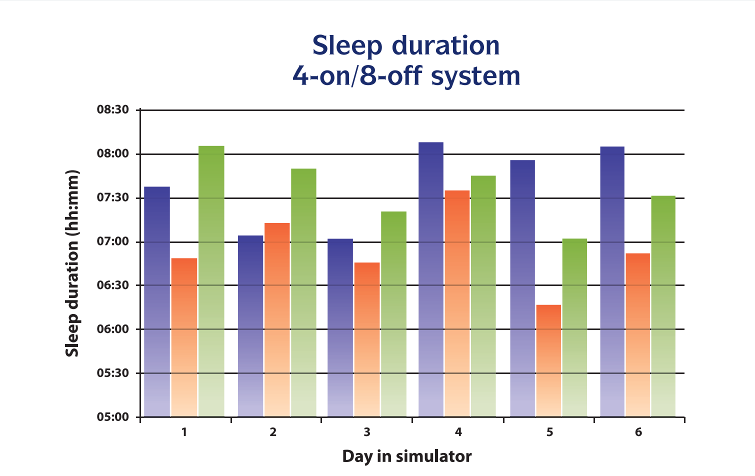


*Fig.6 Actigraphy-based sleep duration per 24-hour period during the experimental week for Chalmers participants working the 4-on/8-off system bridge simulators. The team working 0000 to 0400 and 1200 to 1600 watches is indicated in **blue**; the team working 0400 to 0800 and 1600 to 2000 in **orange**; and the team working 0800 to 1200 and 2000 to 0000 in **green***

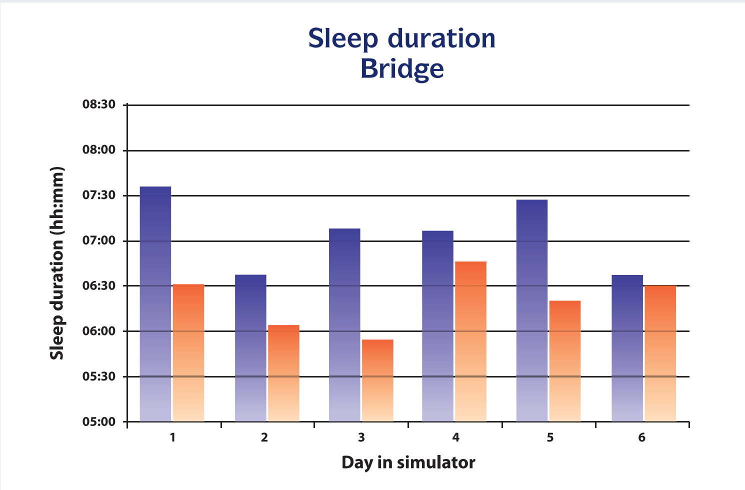


*Fig.7 Sleep duration per 24-hour period during the experimental week for Warsash participants working the 6-on/6-off system bridge simulators. The team working 0000 to 0600 and 1200 to 1800 is indicated in **blue**, the team working 0600 to 1200 and 1800 to 0000 in **orange***

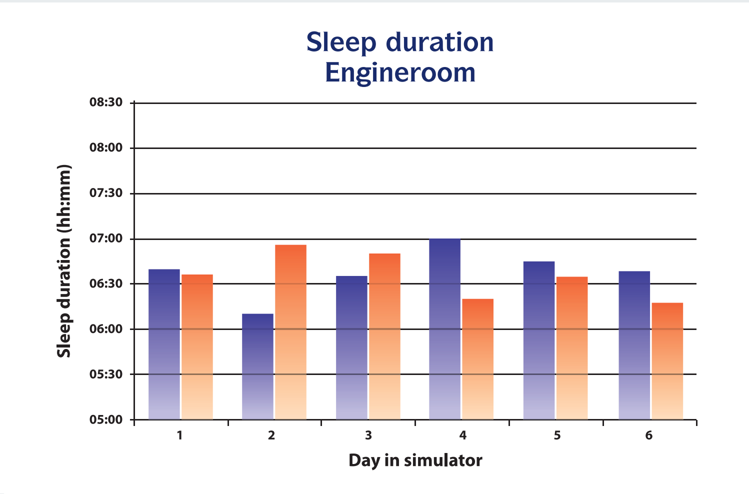


*Fig.8 Sleep duration per 24-hour period during the experimental week for Warsash participants working the 6-on/6-off system engineroom simulators. The team working 0000 to 0600 and 1200 to 1800 is indicated in **blue**, the team working 0600 to 1200 and 1800 to 0000 in **orange***

- **how deep was your sleep? 5 (very deep) to 1 (very light)**
- **did you wake up well rested? 5 (completely) to 1 (not at all)**
- **did you feel stressed at bedtime? 5 (not at all) to 1 (very)**
- **have you slept long enough? 5 (definitely enough) to 1 (definitely too little)**

Total daily sleep duration

The total daily sleep duration for each participant was calculated from the start of the first watch on for every 24-hour period. Variations were found between the two watch systems, with sleep duration being considerably longer in the 4-on/8-off system (211 ±8 minutes per free watch) compared with the 6-on/6-off system (162 ±10 minutes per free watch).

The vast majority of participants were found to split their sleep across the two free watches. In the 6-on/6-off system, sleep duration was longest during the free watches from 0000 to 0600 and from 0600 to 1200. In the 4-on/8-off system, sleep duration was longest during the free watches from 0400 to 1200, from 2000 to 0400 and from 0000 to 0800.

Significant differences were also found in the time taken by participants to get to sleep. The delay to bedtime differed significantly across watches – for example, on Chalmers 4-on/8-off: after 0000-0400 the delay to bedtime was 42 ±21 minutes (±standard error), after 0400-0800 it was 60 ±22 minutes, after 0800-1200 it was 225 ±23 minutes (and few sleeping), after 1200-1600 it was 237 ±18 minutes, after 1600-2000 it was 169 ±20 minutes, after 2000-2400 it was 74 ±23 minutes. Night watches had the least delay to bedtime.

Similarly, at Warsash, a comparison between the bedtimes of deck and engineroom officers on 6-on/6-off (see Figures 9 to 12 on page 20) reveals that after the 0000-0400 watch, the delay to bedtime was generally about 50 minutes, rising to nearer 100 minutes for the late afternoon and evening watches. These results also show that the delay in getting to sleep after the end of the watch was considerably less on the 6-on/6-off regime than for the 4-on/8-off. This relative ease of falling asleep after the end of the watch on 6-on/6-off is explained by the general lack of sleep for those on that more arduous watch.

These results mean that estimates of bed timing need to be adjusted in any estimates of fatigue in models of fatigue regulation.

Sleep on duty

Sleeping on duty poses an immediate and significant threat to safety in any mode of transportation, and shipping is no exception. Through analysis of EEG recordings and visual observation of Project Horizon participants, researchers were able to identify incidents of sleep – both on the bridge and in the engineroom – as defined by the occurrence of at least one 20 second period of stage 1 sleep whilst on watch.

Chalmers 4-on/8-off

The percentage of participants sleeping by watch is indicated in Figure 9. The highest proportion of watchkeepers falling asleep was observed between 0000 and 0400hrs – 40%, or four participants.

McNemar's testing did not reveal any statistically significant differences between watches overall, but between night watches and evening watches. The presence of sleep during day watches (between 1200 and 2000) is unusual and normally not seen, but could be a consequence of working night watches, preventing participants from getting their sleep at the proper time – in the hours of darkness.

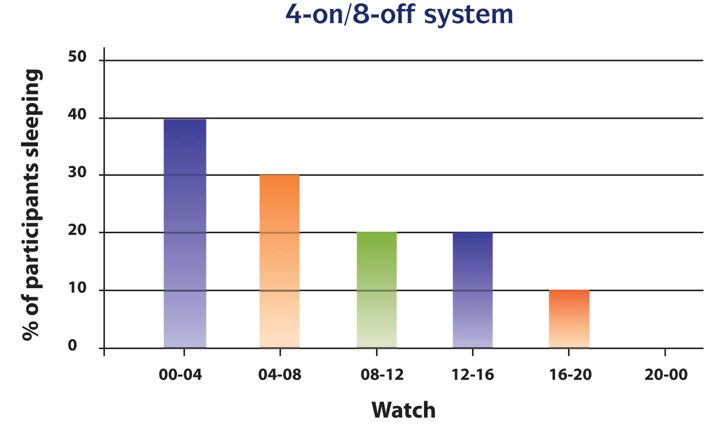


*Fig.9 Percentage of participants sleeping per watch in the Chalmers 4-on/8-off simulations. The team working 0000 to 0400 and 1200 to 1600 watches is indicated in **blue**; the team working 0400 to 0800 and 1600 to 2000 in **orange**; and the team working 0800 to 1200 and 2000 to 0000 in **green***

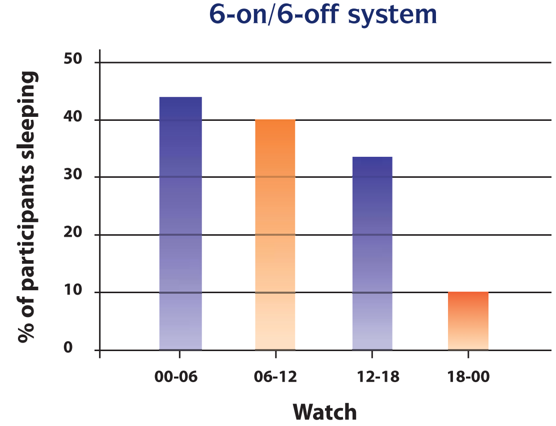


Fig.10 Percentage of participants sleeping per watch in the Warsash 6-on/6-off engineroom simulations

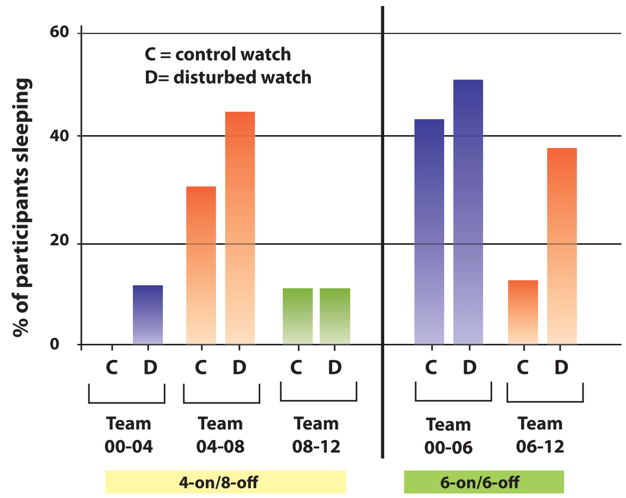


*Fig.11 Percentage of participants sleeping per watch team in the Chalmers simulations after the control watch **(C)** and following the free watch disturbance **(D)**. The teams working in the 4-on/8-off system are shown at the left, those working in the 6-on/6-off system at the right*

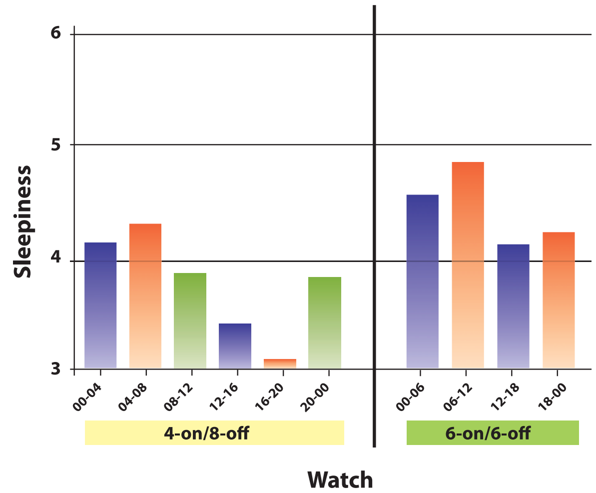


Fig.12 Percentage of participants sleeping per watch team in the Chalmers simulations. The teams working in the 4-on/8-off system are shown at the left, those working in the 6-on/6-off system at the right

Chalmers 6-on/6-off

The percentage of participants sleeping by watch is indicated in Figure 10. The highest proportion of watchkeepers falling asleep was observed between 0000 and 0400 (more than 40%).

Effect of the off-watch disturbance

Increased rates of participants sleeping were noted in almost all watch teams in both watch systems during the watch following the off-watch disturbance (D) as compared to the control watch in the other half of the week (C) – as indicated in Figure 11.

4-on/8-off versus 6-on/6-off

The results (Figure 12) showed more participants sleeping on watch in the 6-on/6-off system than in the 4-on/8-off system, although a level of statistical significance was not reached. At least 50% of participants in both watch teams in the 6-on/6-off system were found to have slept on the bridge, whereas in the 4-on/8-off system such a percentage was only reached by team 1 (working 0000 to 0400 and 1200 to 1600).

Warsash bridge

The percentage of participants sleeping by watch is indicated in Figure 13. The highest proportion of watch keepers falling asleep was observed between 1800 and 0000 (more than 20%, or two participants). McNemar's testing did not reveal any statistically significant differences between watches. Researchers suggest that the absence of participants sleeping between 0600 and 1200 may have been the consequence of half of the group participating in cargo-handling simulations (which is rather activating) at that time. Again, sleep during daytime might have been a consequence of the night work involved.

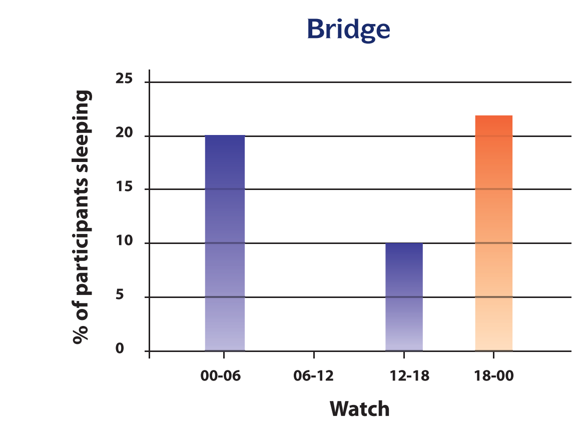


Fig.13 Percentage of participants sleeping per watch in the Warsash 6-on/6-off bridge simulations

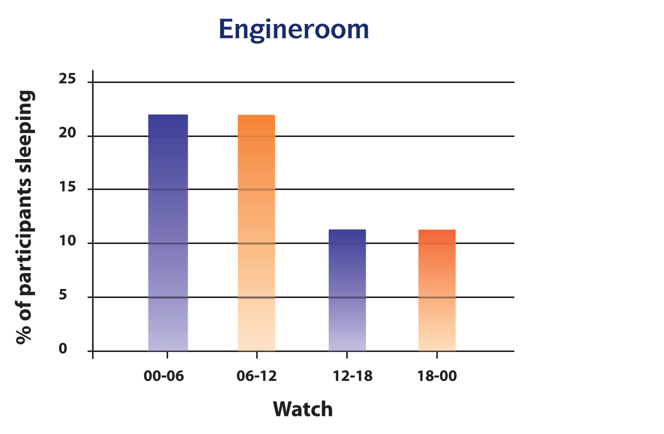


Fig.14 Percentage of participants sleeping per watch in the Warsash 6-on/6-off engineroom simulations

Warsash engineroom

The percentage of participants sleeping by watch is indicated in Figure 14. The highest proportion of watch keepers falling asleep was observed between 0000 and 0600 and between 0600 and 1200 (more than 20%, or two participants).

Bridge versus engineroom

The percentage of participants sleeping on watch was found to be relatively similar for both watch teams in the bridge and the engineroom, as shown in Figure 14. No statistically significant differences were observed.

Conclusions

- **the percentage of participants showing sleep while working on the bridge were unexpectedly high**
- **more participants fell asleep during the night/morning watches than day-early evening watches**
- **a disturbed off-watch period was found to result in more sleep during the subsequent watch**
- **more sleep was found to occur on watch in the 6-on/6-off system than in the 4-on/8-off system**
- **no significant differences were observed between the bridge and the engineroom**

Activity/Electrophysiological measurements

Chalmers 4-on/8-off

Reaction time

Within subjects, the mean reaction time differed across the days and between the first and the second watch of the day, being slower during the first watch. Mean reaction times also differed based on time in watch, being slower at the end of the watch than the start.

Lapses

Within subjects, the number of lapses was greater during the first watch of the day than the second and lapses were more abundant at the end of the watch than at the start.

Chalmers 6-on/6-off

Reaction time

Within subjects, the mean reaction time was found to be slower at the end of the watch than at the start, and the number of lapses was greater at the end of the watch than at the start.

Effect of the off-watch disturbance

In both watch systems, reaction times were slower following the off-watch disturbance. In the 4-on/8-off system, the mean reaction time was considerably slower following the disturbance (306 ±7ms) compared with the control watch (283 ±5ms). The number of lapses was also higher following the disturbance (2.3 ±0.4) compared with (0.9 ±0.2) in the control watch.

In the 6-on/6-off system, no differences in the rate of lapses were observed between subjects following the free watch disturbance and the control watch. However, the mean reaction time was slower following the disturbance: (339 ±27ms) against (289 ±18ms) for the control watch.

4-on/8-off versus 6-on/6-off

Reaction times and number of lapses did not differ between the two watch systems.

Warsash deck

Within subjects, no significant main effects were observed for reaction times or the rate of lapses.

Warsash engineroom

Reaction time

Within subjects, the mean reaction time was found to be slower during the first watch of the day than the second: (339 ±17ms versus 329 ±17ms).

Lapses

Within subjects, the number of lapses was found to be more abundant during the first watch of the day compared to the second: 6.1 ±1.6 versus 5.2 ±1.4. The number of lapses was also more abundant at the end of the watch than at the start: 6.1 ±1.5 versus 5.2 ±1.5.

Deck versus engineroom

PVT reaction times and number of lapses did not differ between the bridge and the engineroom.

Overall findings:

- **worse PVT performance during the first watch of the day**
- **worse PVT performance at the end of the watch compared with the start**
- **the off-watch disturbance worsened PVT performance**
- **PVT performance did not differ between the bridge and the engineroom**

Stroop test

Warsash deck

The reaction time on control stimuli did not differ within or between subjects. However, within subjects, the mean reaction time on interference stimuli differed significantly across days with a gradual decline in daily means (1103 ±61ms on day 1 to 982 ±58ms on day 7) which indicated a learning effect over the course of the week.

The number of mistakes on control stimuli did not differ within or between subjects and no mistakes on interference stimuli were observed within subjects. However, the number of mistakes on interference stimuli differed significantly between the two watch teams, with the team working 00:00-06:00 making more mistakes (2.1 ±0.4) than the other team (0.6 ±0.3).

Within subjects, absolute interference (the mean reaction time on interference stimuli minus the mean reaction time on control stimuli) differed significantly across days, with the gradual decline in daily means (from 136 ±18ms on day 1 to 60 ±20ms on day 7) suggesting the presence of a learning effect. Between subjects, no effects were observed.

Within subjects, percentual interference (the relative increase in reaction time on interference stimuli as compared to control stimuli) differed significantly across days and the gradually declining daily means (from 13.8 ±1.7% on day 1 to 6.3 ±2.0% on day 7), were a sign of a learning effect. Between subjects, no effects were observed.

Warsash engineroom

Within subjects, the mean reaction time on control stimuli differed significantly across days and the gradual decline in daily means (909 ±51ms on day 1 to 805 ±37ms on day 7) indicated a learning effect over the course of the week. The mean reaction time was observed to have differed between the first and the second watch of the day, with slowermean reaction times during the first watch (857 ±47ms) than during the second (832 ±41ms). Between subjects, no effects were observed.

Within subjects, the mean reaction time on interference stimuli differed significantly across days, with the gradual decline in daily means (1010 ±69ms on day 1 to 883 ±54ms on day 7) being indicative of a learning effect over the course of the week.

The number of mistakes on control stimuli and interference stimuli did not differ within or between subjects. Within subjects, absolute interference (the mean reaction time on interference stimuli minus the mean reaction time on control stimuli) differed significantly across days and the gradual decline in daily means (from 102 ±24ms on day 1 to 79 ±24ms on day 7) suggested the presence of a learning effect. Between subjects, no effects were observed. Percentual interference did not differ within or between subjects. Stroop performance was not found to differ depending on the watch.

Conclusion

Deck versus engineroom

None of the Stroop test parameters differed between the bridge and the engineroom. Overall, the tests showed slower reaction times on interference stimuli than on control stimuli. This interference effect declined during the course of the week, probably due to a learning effect.

A participant wired up for Actigraph and EEG recordings during the bridge simulations

Naturalistic performance

Chalmers 4-on/8-off

Subjects were observed responding to a range of 'events' and sub-tasks, including:

- **collision course**
- **compliance to collision regulations**
- **the presence of fishing boats**
- **object adrift**
- **communication task**
- **close encounter**
- **high-speed ferry**

Whole watch performance was assessed and if nodding-off was observed by simulator instructors and/or researchers, the watch was scored as 1. If not, it was scored as zero. No differences within or between subjects were observed.

Performance during the two sessions in the liquid cargo operations simulators was expressed on a 0 to 100 scale. Performance did not differ within or between subjects.

Comparative analysis was not possible for a number of events and sub-tasks but, overall and in particular in relation to the 'close encounter' event, a limited performance increase was noted during the course of the week and this was indicative of a learning effect.

Chalmers 6-on/6-off

Comparative analysis was not possible for a number of events and sub-tasks but, overall performance in the 'close encounter' event was higher in the team working 00:00 to 06:00 watch (7.1 ±0.5 versus 3.8 ±0.5).

No differences within or between subjects were observed in relation to nodding-off or performance during the two sessions in the liquid cargo operations simulators.

Effect of the free watch disturbance

Limited and somewhat bi-directional effects were observed. The off-watch disturbance only affected sub-task 1 on the whole watch performance rating. In the 4-on/8-off system, a significant difference was observed on this sub-task between the control watch and the watch following the off-watch disturbance, with performance being worse following the disturbed off-watch period (2.8 ±0.2) than in the control watch (3.2 ±0.2). This effect was not seen in the 6-on/6-off system, but an interaction between day and watch team was observed.

Comparing the free watch following the disturbance with the control watch, performance increased in the team working 00:00 to 06:00, but decreased in the other team.

In the 4-on/8-off system, more nodding offs were observed after the free watch disturbance than in the control watch: 0.3 ±0.1, against none during the control watch.

Monitoring bridge operations during a simulated voyage

4-on/8-off versus 6-on/6-off

Limited differences between the two watch systems were noted from all sub-tasks of all events. Performance on sub-task 3 (detection range) of event 3 (communication event) was seen to be higher in the 6-on/6-off system (4.0 ±0.3 versus 2.2 ±0.2). Sub=task 1 (position taking) of the whole watch performance was found to be higher in the 4-on/8-off system (2.9 ±0.1 versus 2.6 ±0.1).

Warsash deck

Adherence to the collision prevention regulations was scored during the first and second watch of days 2, 4, and 7 and rated on a 0 to 10 scale. No within or between subjects effects were observed.

The standard of communications with the engineroom was measured in the second watch on days 4 and 7. Limited differences were noted and only accuracy differed between the days, with higher accuracy being observed on day 4 than on day 7 (4.8 ±0.2 versus 4.1 ±0.3). The effect of day was only present in the team working 00:00 to 06:00.

Accuracy and completeness of the watch handover was rated on a 0 to 10 scale during both watches on days 2, 4, and 7. No within or between subjects effects were observed.

Overall log keeping was rated on a 1 to 10 scale on days 2, 4, and 7 during both watches and was found to be rated higher during the second watch of the day than the first (6.3 ±0.6 versus 5.8 ±0.6).

No within or between subject effects were observed in standard alteration of course and determination of position tasks.

However, several LICOS performance scores decreased during the course of the week, including: adherence to standard watchkeeping tasks (from 4.5 ±0.3 on day 1 to 4.2 ±0.2 on day 3 to 3.7 ±0.3 on day 6) and completeness of the handover (from 4.4 ±0.3 on day 1 to 4.3 ±0.3 on day 3 to 3.3 ±0.3 on day 6). The decline in the completeness of the watch handover over the three days was much more pronounced in the team working 00:00 to 06:00 (from 5.2 ±0.5 to 3.1 ±0.4 versus 3.5 ±0.5 to 3.5 ±0.4).

Warsash engineroom

No differences between or within subjects were noted on a range of tasks, including professional discussion, providing current status information and acknowledgement of information received.

The team working the 00:00 to 06:00 watch scored higher in requesting information when coming on watch (4.1 ±0.2 versus 3.5 ±0.2) and higher on one of the problem-solving tasks(2.5 ±0.3 versus 1.6 ±0.3) when responding to a high scavenge air temperature on the main engine.

Performance ratings for a range of speed and accuracy tasks varied – with some decreasing during the week, others improving and no differences within or between subjects being noted on others.

Adherence to standing orders was found to be greater in the second watch of the day than the first (4.8 ±0.5 versus 4.6 ±0.1). The quality of the watch handover was found to have gone from 6.7 ±0.4 on day 3 to 7.8 ±0.2 on day 5 and overall watch performance ranged from 6.8 ±0.4 on day 3 to 7.7 ±0.2 on day 5.

Overall, the team working 00:00 to 06:00 performed slightly better. Performance on some tasks increased during the course of the week, while performance on other tasks decreased.

Conclusions

Project Horizon has undoubtedly succeeded in its core of aim of delivering a more informed and scientifically rigorous understanding of the way different watchkeeping patterns at sea affect the performance of ships' officers. The range of measurements and the high degree of realism gained through the use of simulators have provided detailed and robust data on which to assess and analyse effects. Data gained from the research is sufficiently robust to provide input to marine-validated mathematical fatigue prediction models within a fatigue risk management system.

Overall, it is clear that much of the data gained from the research supports the 'circadian theory' of diurnal performance peaks and troughs and clear evidence of 'sleepiness' risk periods.

- **watchkeepers were found to be most tired at night and in the afternoon**
- **sleepiness levels were found to peak towards the end of night watches**
- **slowest reaction times were found at the end of night watches**
- **incidents of sleep on watch mainly occurred during night and early morning watches**
- **the 6-on/6-off regime was found to be more tiring than 4-on/8-off**
- **the onset of tiredness on 6-on/6-off occured over a shorter timeframe than predicted**
- **'disturbed' off-watch periods produce significantly high levels of tiredness**
- **participants on 6-on/6-off rotas were found to get markedly less sleep than those on 4-on/8-off**
- **all groups reported relatively high levels of subjective sleepiness on the KSS scale**

EEG data demonstrated that a large proportion of the watchkeepers showed actual sleep on the bridge, particularly following 'disturbed off-watch' periods. In the Chalmers 4-on/8-off pattern, the figures varied between 0% for the 1200-1600 and 2000-0000 watches to almost 40% on the 0000-0400 watch. In the Chalmers 6-on/6-off rotas, the figures varied between almost 10% for the 1800 to 0000 watch to more than 40% for the 0000-0600 period. The disturbed off-watch period was also found to have a strong impact on sleepiness.

Total daily sleep duration was measured for all participants over each 24-hour period and marked differences were detected between the different watchkeeping patterns. The differences were particularly apparent amongst those working the 6-on/6-off schedules – where data showed less sleep and a clear 'split' sleeping pattern in which daily sleep was divided into two periods, one of between three to four hours and the other averaging between two to three hours. In contrast, the sleep patterns for those working the 4-on/8-off schedule at Chalmers were found to be relatively normal – varying between around 7 to 7.5 hours for Team 1 to around 6 hours for the second team.

Another important observation on sleep was that it was found not to be initiated almost immediately after ending a night watch, but there was a delay averaging about 50 minutes. This is important information for estimating the recuperative effect of sleep. Off-watch periods starting in the afternoon to evening were found to have very long delays to the onset of sleep and often no sleep was taken, particularly on the 4-on/8-off pattern. See Figure 15 opposite.

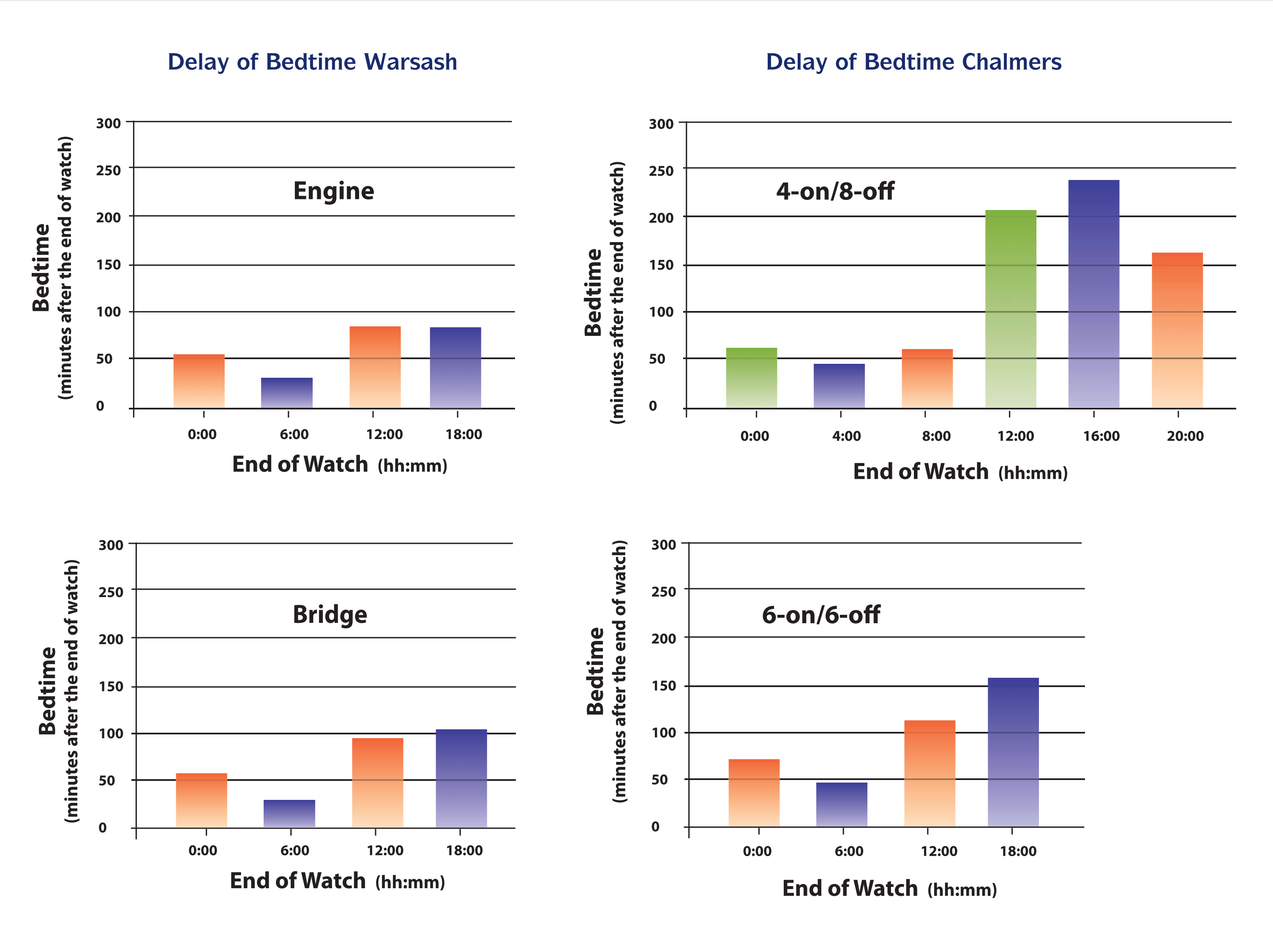


Fig.15 An important finding from Project Horizon was data showing how watchkeepers can experience very long delays to the onset of sleep following their watch

Outcomes

There can be no doubt that Project Horizon has achieved its principal objective of gaining a deeper and more scientifically rigorous understanding of the way in which sleepiness affects watchkeepers at sea. The results have taken knowledge of the issues to a new level and have demonstrated the multiple and complex effects of some of the most common working patterns for seafarers.

It should be noted, however, that Project Horizon was a simulator-based study that was designed to study some basic aspects of the effects of standard maritime watch schedules on sleepiness and fatigue. Whilst the simulator setting can present a limitation, it does provide better control of the test conditions and offered researchers opportunities for in-depth comparative analysis of participants at different times and on different working patterns in near-identical situations. Whilst every effort was made to design realistic simulated working conditions, the practical limitations must be recognised – such as timescales and working environment. There are many other factors may have an important impact on watchkeepers' sleep and rest – such as bad weather conditions, onboard noise, the effects of long periods at sea, skills and competence of the crew, and varying rules on the use of chairs on the bridge. All these are influences that need to be considered in future studies of fatigue at sea.

Nevertheless, Project Horizon has delivered an unprecedented level of remarkably detailed data that enables the achievement of the core objective of using the findings to assist the development of 'best practice' standards for the shipping industry. The results also provide reliable and validated source material for input into policy discussions at national, regional and international level – with the potential for appropriate bodies to take forward plans for improved regulation of seafarers' working hours, safe manning and fatigue mitigation.

Analysis and assessment of this data has enabled researchers to develop a lasting legacy, in the form of a proposed fatigue management toolkit. This package is intended to provide practical guidance for key stakeholders covering:

- **the nature of fatigue or sleepiness at sea**
- **pointers to aid recognition of such conditions**
- **measures by which mitigation of them might be achieved**
- **concrete indications how the conditions might be avoided at source and the findings of the project might be applied – in particular to the key stakeholders: seafarers; ship owners/managers; classification societies; policymakers/ regulatory authorities; training establishments; equipment providers**

Fatigue management toolkit

Sleepiness is an acknowledged risk factor in safety-critical industries and in all modes of transport. It is recognised, however, that shipping differs from some other transport modes, in that the nature of risk exposure and the capacity to act is extremely variable and depends on many factors. The characteristics of working at sea – and especially in the deepsea trades – mean that the coincidence of exposure to risk and absence of capacity to deal with it will be a relatively rare event. It is probable that the level of risk will be much lower than that for road transport, for example, and most likely to be more similar to that in aviation.

In fact, the data from Project Horizon indicates that the probability of danger at sea will be highest when night watches are combined with prior reduction of sleep opportunities, and exacerbated by passages through narrow or very densely travelled waters, or during reduced visibility.

The Project Horizon findings suggest that owners, regulators, seafarers and others should pay special attention to the potential risks in difficult waters in combination with the 6-on/6-off watch system (because of sleep loss), night watches, the last portion of most watches (especially night watches), and watches after reduced sleep opportunity. There is also some evidence from the research to suggest that individual susceptibility to fatigue probably also needs to be considered.

A variety of methods (some of which are already commonly deployed) may be used to address this potential risk, including alarm systems to alert crew before important waypoints, encouragement not to use chairs on the bridge during night watches, additional crew, training crew to recognise symptoms of fatigue, and special protection of sleep periods for watchkeepers.

Another way of reducing fatigue-related risk is to train seafarers in understanding the causes and consequences of fatigue, how to detect it, how to prevent it and how to report it. The latter requires a level of acceptance of fatigue reporting without reprisals from those in authority. Personal fatigue countermeasures include caffeine, strategic napping and physical or mental activity. Judicial use of countermeasures against fatigue should be part of the job description for all personnel on watch duty.

The toolkit takes these precautions a step further, by using scientifically verified data to build mathematical models which can be used to predict which portions of a particular voyage may be critical from a fatigue point of view – allowing mitigating action to be planned ahead of time.

It is well known that working hours which deviate from conventional patterns (shift work, roster work, and irregular watch schedules) always entail a high probability of reduced sleep and of increased fatigue, with an ensuing accident risk. In recent years, scientists have developed mathematical models for alertness or performance prediction – and these have most notably been applied in the aviation industry. An example of the recognition of the value of such systems can be seen from the US National Transportation Safety Board's 'Most Wanted List' and the associated 2011 recommendation stating: 'The Safety Board continues to call for the development of fatigue management systems, which take a comprehensive approach to reducing fatigue-related risk. These systems should be based on empirical and scientific evidence and should include a methodology to continually assess their effectiveness.'

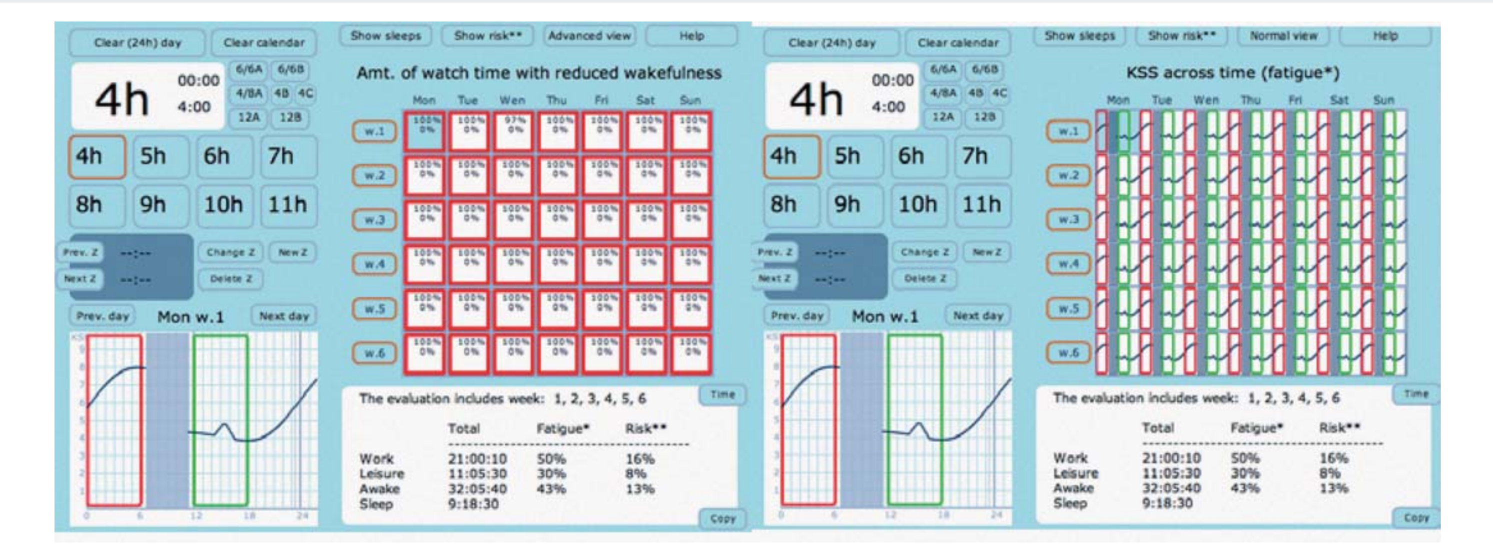
It is against this background that the Project Horizon researchers have been able to use the robustness of the results of their work to develop a maritime alertness regulation version of these models – '**MARTHA**': an acronym derived from 'a maritime alertness' regulation tool based on hours of work.

Mathematical models for alertness or performance prediction have been developed mainly as tools for evaluating the effects on sleepiness or fatigue of work schedules or sleep/wake patterns that deviate from the pattern of daytime activity and night time sleep. Early models were based on the effects of time awake and amount of prior sleep as well as a circadian component representing the effect of the biological clock. As scientific understanding has increased, models have become more sophisticated, incorporating a wider range of factors that influence sleepiness and alertness and expanding to include predictions of sleep latency and sleep duration.

The detailed information obtained in Project Horizon has enabled the model to be validated against the empirical sleepiness data. Apparently, there has been no prior knowledge of the way in which sleep is distributed across sleep opportunities on sea schedules. An important new development from Project Horizon has been the use of the empirical sleep data (bedtimes and rise times) obtained from the research to create a new function of the model to predict sleep on sea watches. The model also incorporates a third process reflecting the effects of time on-watch.

These functions were combined and, using a computer-based system, will provide a maritime interface with selectable watch schedules and a 'do-it-yourself' watch system facility. Users will be able to enter their working schedules over a six-week time window and receive predicted estimates of the most risky times and the times of highest potential sleepiness for each watch and for the whole watch schedule, as well as for time outside watch duty.

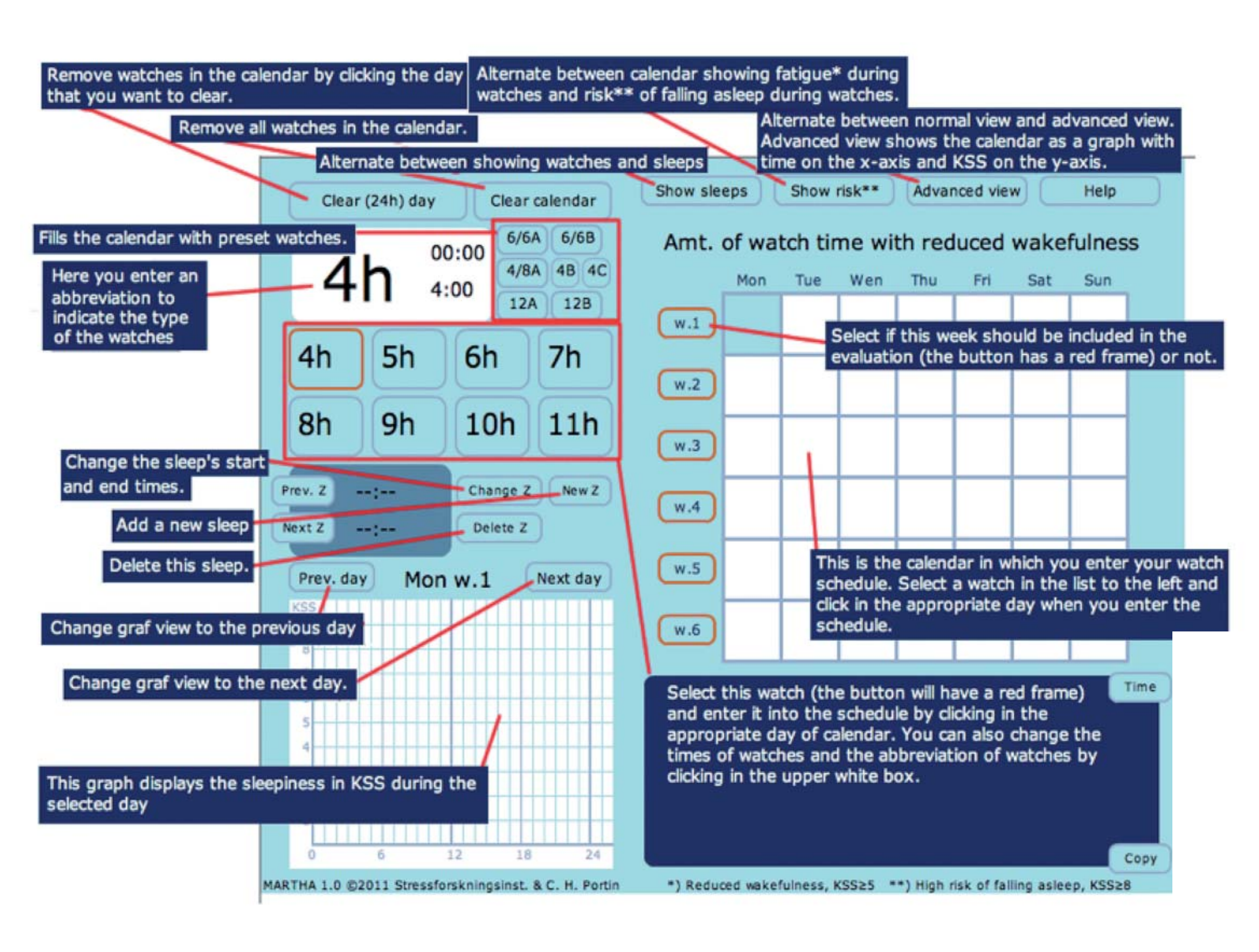
The major display contains estimates for each 24-hour period, with a second display to describe each 24-hour period with sleep periods and a continuous estimate of sleepiness. This information may also be displayed as miniatures in the main display.



How MARTHA could predict sleepiness on a 6-on/6-off schedule for team A (0-6 +12-18). 16% of the time on watch sleepiness is at dangerous levels. The second image includes the miniatures of the continuous curve and the predicted sleep periods

MARTHA could be used onboard during voyage planning to develop watch systems that are efficient and that minimise risk. Shipping companies can use the system when planning the size and competence of the crew. The tool could also yield important International Safety Management Code benefits, and might be used for insurance and classification purposes.

MARTHA could also assist flag states and port state control authorities, enabling solid documentation if, for example, a ship is to be detained in order to let the crew rest before the voyage is resumed. It could also be used for the prevention and investigation of accidents. The Horizon consortium recognises that Project Horizon is a project that will be more for public benefit than having commercially exploitable outputs. It fulfils a need that could not economically be sustained by any individual, or even a group of, actors, without the essential ingredient of public funding, by courtesy of the EU. It will achieve its success through exploitation, in a variety of ways: widely and generally on the world stage of maritime safety; as well as individually through the benefits attained and appreciated by the project partners themselves.



The MARTHA interface

Recommendations

The overall results from Project Horizon may be transferred into different types of recommendations. However, these need to acknowledge the total risk situation – the convergence of risk exposure and capacity to act. In road transport the risk exposure is present 100% of the time. In seafaring risk exposure may mainly occur in manoeuvring in narrow or otherwise difficult waters or with poor visibility. The incidence of such exposure will vary greatly depending on many factors, but must be very much lower than that for road transport – probably more similar to that in aviation.

One of the strongest factors influencing the capacity to act is sleep, when performance is absent. However, such states during work are relatively sparse and sporadic – even during night work – but they occur for most operators on each difficult watch or shift.

The coincidence of exposure to risk and absence of capacity to deal with it will be a relatively rare event. The probability of danger will be highest when night watches are combined with prior reduction of sleep opportunities, together with passages through narrow or very densely travelled waters, or during reduced visibility.

Considering the results of the present study, special attention needs to be paid to:

- **the risks in passages through difficult waters in combination with the 6-on/6-off watch system (because of sleep loss)**
- **night watches**
- **the last portion of most watches (especially night watches)**
- **watches after reduced sleep opportunity**
- **individual susceptibility to fatigue also needs to be considered**

The suggested 'special attention' may involve alarm systems to alert crew before important changes of course, alerting devices, encouragement not to use chairs on the bridge during night watches, additional crew, special protection of sleep periods for watchkeepers, or no work apart from watchkeeping.

In addition, mathematical models (MARTHA) can be used to predict which portions of a particular voyage may be critical from a fatigue point of view and thereby mitigating action can be planned ahead of time.

One way of reducing risks related to fatigue may also be to train the crew in the causes and prediction of fatigue, its risks, how to detect it, how to prevent it and how to report it. The latter requires a level of acceptance of fatigue report without reprisals for those in authority. Personal fatigue countermeasures include caffeine, strategic napping and physical or mental activity. Judicial use of countermeasures against fatigue should be part of the job description for all personnel on watch duty.

Most of the general points discussed above are part of what is called 'Fatigue Risk Management', and which is presently being implemented in aviation worldwide. A similar development seems called for in marine operations.

A final recommendation concerns future research. Project Horizon is the first detailed and experimental study of fatigue at sea. As discussed previously, it has limitations, one of which is that the data has been obtained in a simulator. This makes good experimental control possible, but also detracts from the possibility to generalise. There is a clear need for replicating the present study at sea and to carry out studies of long periods at sea to identify fatigue causes that may derive from boredom, isolation and similar factors.

Worst case scenario

This was one of the worst cases observed in Project Horizon. It involved one of the Warsash participants and occurred on day 2, during his watch from 1800 to 0000. Physiological sleepiness during the course of this watch is shown in Figure 16, where sleepiness ranges from stage 0 (fully alert) to stage 100 (fully asleep). It can be seen that after less than 1.5 hours in watch, the participant started to show rapid alternations between wakefulness and sleep. Of special concern is the period between about 2005 and 2035, where the participant was in a state of continuous sleep. Such a sleep period occurred again between about 2245 and 2300. Figure 17 shows the EEG trace at 2029. The high amplitude low frequency waves on the upper two traces indicate a state of deep sleep. Deep sleep on duty is a very serious condition, because it requires additional effort to get woken up from than from lighter sleep. In addition, waking up straight from deep sleep is usually followed by a state called sleep inertia from which a substantial amount of time after waking up will be required to regain full alertness again.

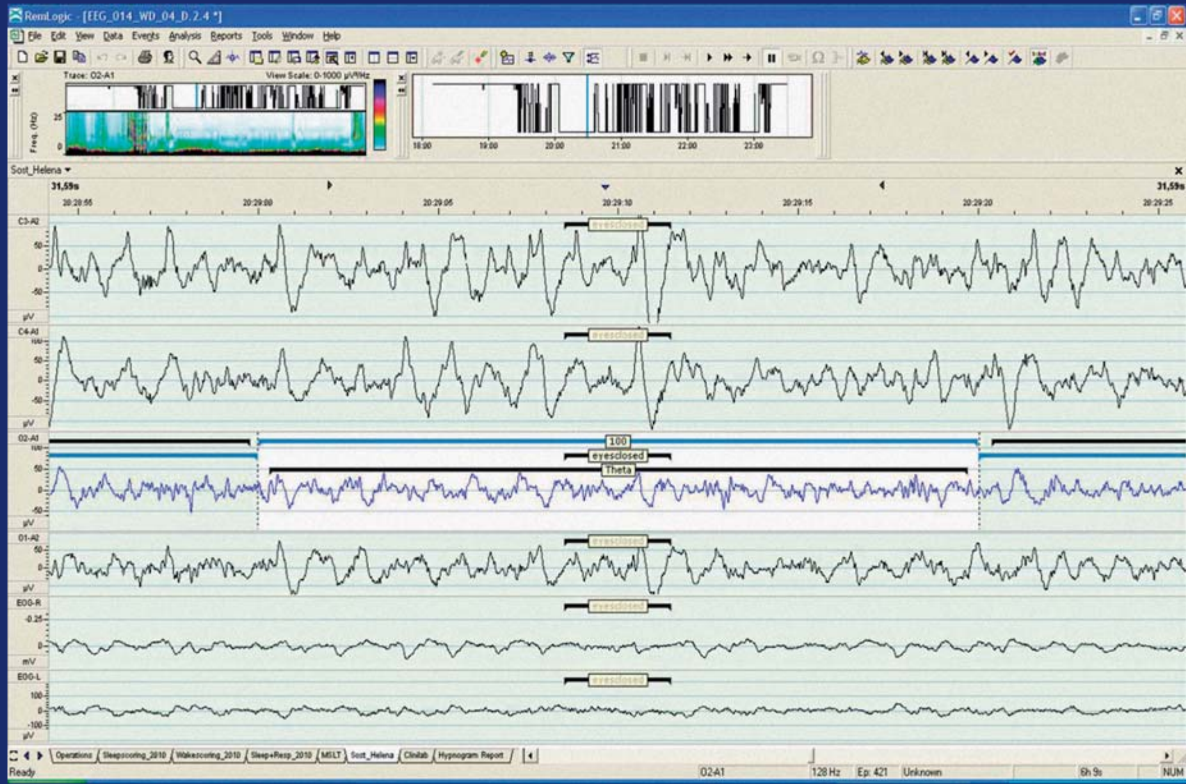


Fig.16

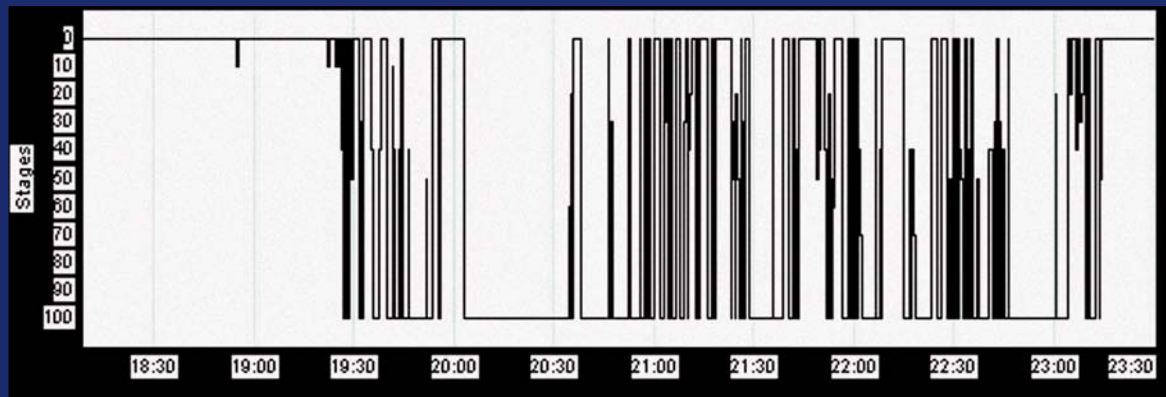


Fig.17

www.project-horizon.eu
Email: graham.clarke@inchmery.eu

Project Horizon consortium

Southampton Solent University – Warsash Maritime Academy (Coordinator) (WMA)
Bureau Veritas – Marine Division, Research Dept (BV)
Chalmers Tekniska Hoegskola AB – Dept of Shipping & Marine Technology (Chalmers)
European Transport Workers' Federation – Nautilus International (ETFN)
Stockholms Universitet – Stress Research Institute (SU)
The Standard P&I Club – Charles Taylor & Co Limited (CTPI)
European Community Shipowners Associations (ECSA)
European Harbour Masters Committee (EHMC)
International Association of Independent Tanker Owners – Intertanko (INTKO)
UK Marine Accident Investigation Branch (MAIB)
UK Maritime & Coastguard Agency (MCA)