Aus Institut und Poliklinik für Arbeits-, Sozial- und Umweltmedizin

Klinikum der Ludwig-Maximilians-Universität München

Direktor: Prof. Dr. med. Dennis Nowak

Flight safety in helicopter emergency medical services:

An investigation of potential hazards with special consideration of pilot age

Dissertation

zum Erwerb des Doktorgrades der Humanbiologie

an der Medizinischen Fakultät der

Ludwig-Maximilians-Universität zu München

vorgelegt von

Hans Peter Werner Bauer

aus

Weilheim i. OB

2019

Mit Genehmigung der Medizinischen Fakultät

der Universität München

Berichterstatter:	PD Dr. Britta Herbig
Mitberichterstatter:	Prof. Dr. Bernhard Zwißler
	Prof. Dr. Matthias Graw
	Prof. Dr. Christian K. Lackner
Dekan:	Prof. Dr. med. dent. Reinhard Hickel
Tag der mündlichen Prüfung:	15.07.2019

Eidesstattliche Versicherung

Bauer, Hans

Name, Vorname

Ich erkläre hiermit an Eides statt,

dass ich die vorliegende Dissertation mit dem Thema

> Flight safety in helicopter emergency medical services: An investigation of potential hazards with special consideration of pilot age

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, 25.07.2019	Hans Bauer
Ort, Datum	Unterschrift Doktorand

Danksagung

Mein allererster, herzlicher Dank gilt PD Dr. Britta Herbig für die stets offene Tür und das offene Ohr, für die immer angenehme und konstruktive Zusammenarbeit und die vielen anregenden Diskussionen (ob nun über die Wissenschaft, den Sinn des Lebens oder das Kantinenessen).

Aus denselben Gründen (na gut, vielleicht abgesehen von manchen der Diskussionsthemen) möchte ich mich auch ganz herzlich bei Prof. Dennis Nowak bedanken.

Danke an alle Studienteilnehmer sowie an die vielen Helfer bei den teilnehmenden Luftrettungsdiensten, flugmedizinischen Instituten und fliegerärztlichen Praxen für das entgegengebrachte Vertrauen, die Unterstützung und die gute Zusammenarbeit.

Tausend Dank allen, die mir beim Verfassen dieser Arbeit geholfen haben, einfach indem sie da waren.

Table of contents

Abbreviations

AeMC	Aeromedical center
AME	Aeromedical examiner
BMI	Body mass index
CAC	Coronary artery calcium
CFR	(United States) Code of Federal Regulations
EASA	European Aviation Safety Agency
EHAC	European HEMS & Air Ambulance Committee e.V.
EU	European Union
FAA	(United States) Federal Aviation Administration
HEMS	Helicopter emergency medical services
ICAO	International Civil Aviation Organization
IMC	Instrument meteorological conditions
JAA	Joint Aviation Authorities
LMU	Ludwig-Maximilians-Universität
LuftPersV	Verordnung über Luftfahrtpersonal
PIC	Pilot in command
QTc interval	Corrected Q-wave – T-wave interval (electrocardiogram)
SCORE	Systematic Coronary Risk Evaluation (cardiovascular risk score)
TAWS	Terrain awareness and warning system
US	United States
VFR	Visual flight rules

Summary

Introduction

Helicopter emergency medical services (HEMS) play an increasingly important role in many countries' emergency medical care systems. However, whether HEMS provides additional benefit over conventional ground services depends on many factors, including mission safety. In this thesis, results from a study investigating different potential hazards to flight safety in HEMS are reported.

Methods

Data on participating active HEMS pilots working for one of five operators in four European countries were collected in 2015/16 from three sources: (a) flight simulator performance ratings, (b) aeromedical examinations records from the preceding 10 years, and (c) questionnaires completed by the pilots. Medical fitness and psychological fitness were related to potential hazard factors using different analytic approaches.

Results

On the whole, the longitudinal development of the cardiometabolic risk marker profile of pilots aged 60 and over at the end of follow-up was more favorable than that of younger pilots (with large unexplained interindividual differences in risk marker levels). Perceived rewards at work, perceived predictability of work, and physiological dysregulation were selected as the most powerful predictors of simulator performance by a machine learning method from a large array of candidate predictors. Further yet unpublished results include analyses of the relation of pilot age to flight simulator performance and to physiological dysregulation, and of the work stress-strain profile of HEMS pilots.

Conclusions

The results suggest the presence of a healthy worker survivor effect, calling into question the empirical basis of existing European legislation which bans pilots aged 60 and over from conducting single-pilot commercial air transport operations. Application of machine learning methods may open new avenues for identifying flight safety hazard factors. Finally, the results point to the importance of occupational stressors and resources for HEMS flight safety.

Zusammenfassung

Einleitung

Helikopter-Rettungsdienste (helicopter emergency medical services; HEMS) werden in der notfallmedizinischen Versorgung immer wichtiger. Der Zusatznutzen von HEMS im Vergleich zu bodengebundenen Rettungsdiensten ist jedoch von vielen Faktoren abhängig, so auch der Einsatzsicherheit. In dieser Dissertationsschrift werden Ergebnisse einer Studie über potentielle Risiken für die Flugsicherheit bei HEMS berichtet.

Methoden

Über die teilnehmenden, bei einem von fünf Rettungsdiensten aus vier Ländern arbeitenden aktiven HEMS-Piloten wurden 2015/16 Daten aus drei Quellen erhoben: (a) Flugsimulatorleistungsbewertungen, (b) Befunde von fliegerärztlichen Untersuchungen der vergangenen 10 Jahre, und (c) von den Piloten ausgefüllte Fragebögen. Medizinische und psychologische Fitness wurden anhand verschiedener analytischer Methoden zu potentiellen Risikofaktoren in Bezug gesetzt.

Ergebnisse

Die längsschnittliche Entwicklung des kardiometabolen Risikomarkerprofils von Piloten, die zum Ende des Follow-Up 60 Jahre oder älter waren, war insgesamt betrachtet günstiger als bei jüngeren Piloten (bei großen unerklärten interindividuellen Unterschieden in den Ausprägungen der Risikomarker). Wahrgenommene arbeitsbezogene Belohnung und Vorhersagbarkeit der Arbeitsanforderungen sowie physiologische Dysregulation wurden von einem maschinellen Lernverfahren als die leistungsstärksten Prädiktoren aus einer großen Menge von Kandidatenprädiktoren ausgewählt. Weitere, noch unpublizierte Ergebnisse umfassen Analysen des Zusammenhangs des Pilotenalters mit Simulatorleistung und mit physiologischer Dysregulation, sowie des arbeitsbezogenen Belastungs-Beanspruchungs-Profils der HEMS-Piloten.

Schlussfolgerungen

Die Ergebnisse deuten auf einen „Healthy Worker Survivor“-Effekt hin, was die empirische Basis bestehender europäischer Gesetzgebung in Frage stellt, welche gewerblichen Luftverkehrsbetrieb durch Einzelpiloten über 60 Jahren verbietet. Die Anwendung maschineller Lernverfahren könnte neue Wege eröffnen, um flugsicherheitsgefährdende Faktoren zu identifizieren. Außerdem weisen die Ergebnisse auf die Wichtigkeit arbeitsbezogener Stressoren und Ressourcen für die Flugsicherheit von HEMS hin.

1. Introduction

1.1 Utility of helicopter emergency medical services

The importance of helicopter emergency medical services (HEMS) for the transportation of critically ill or injured patients has increased significantly over the previous decades. For instance, in Germany there were about 20,000 to 30,000 missions per year in the 1980s; by the end of the 2000s, this figure had increased to approximately 100,000 missions per year (Hinkelbein, Schwalbe, Neuhaus, Wetsch, & Genzwürker, 2011). Providing life support measures and professional medical care as soon as possible (within the so-called “golden hour”; Cowley, Mergner, Fisher, Jones, & Trump, 1979; Newgard et al., 2010) to patients who suffered severe trauma, burns, cardiovascular events, or other critical conditions was arguably the main motivation for the establishment of HEMS in many developed countries since about the 1970s. Further driving factors included positive experiences with aeromedical use of helicopters in the military, rising numbers of automobile traffic accidents, and the possibility to reach patients in otherwise inaccessible terrain (Galvagno et al., 2015; Kessler, 2015).

At the same time, the use of HEMS has not been, and still is not, without controversy. Already in the early seventies, the United States National Highway Traffic Safety Administration advised against federal funding of HEMS due to an unfavorable cost-to-benefit ratio (US Department of Transportation, 1972). Compared to ground ambulance services, HEMS is often considerably more expensive, e.g. due to higher vehicle, personnel, or transportation costs (Brazier, Nicholl, & Snooks, 1996; de Wing, Curry, Stephenson, Palmieri, & Greenhalgh, 2000). Moreover, determining whether HEMS use is appropriate given the limited available information at the time of an emergency call is notoriously difficult for the dispatcher, and often leads to overtriage; for example, on average between 60 and 70% of trauma patients transported by HEMS from the scene to a hospital were subsequently found to have sustained only minor injuries according to one meta-analysis (Bledsoe, Wesley, Eckstein, Dunn, & O'Keefe, 2006), and 11% of all missions carried out by the largest German civilian HEMS operator in 2016 were to no avail since the patient was not present any more at the scene upon arrival of the helicopter (ADAC Luftrettung, 2017). Finally, a number of studies did not find any survival improvement in HEMS-transported patients compared to ground transport (Bledsoe, 2003).

Yet, according to a recent systematic review of trauma survival in HEMS versus ground transport which was conducted according to Cochrane Collaboration standards (Galvagno et al., 2015), in a large majority of studies, there were positive and often significant effects of HEMS transport on survival after adjustment for differences in patient characteristics, most importantly injury severity. This suggests that despite overtriage, HEMS patients are on average more severely injured and particularly in these cases, HEMS can play an important role in an integrated trauma system (Doucet et al., 2013). The overall very large heterogeneity in the findings of empirical studies on HEMS benefits, costs, and cost-effectiveness (Butler, Anwar, & Willett, 2010; Galvagno et al., 2015; Taylor et al., 2010) is not surprising given the many (interdependent) factors and circumstances involved which may differ between HEMS systems or even HEMS units within a system, such as prevailing geography and population density (Kessler, 2015), geographical distribution of HEMS units (Branas et al., 2005; Brown, Rosengart, Billiar, Peitzman, & Sperry, 2017), differences in crew configuration, available medical equipment and expertise (Butler et al., 2010; Rasmussen, Røislien, & Sollid, 2018), integration of HEMS into the overarching emergency medical services system (Bledsoe, 2003; Habib et al., 2014; Kessler, 2015), quality of technical equipment, availability of appropriately trained flight personnel, and acceptance of HEMS by the local population (Ringburg et al., 2009).

1.2 Flight safety and hazard factors in HEMS

Given, then, that the utility of HEMS crucially depends on system design and the overall balance of such contributing factors, there is a continuing need to optimize aspects of the system while considering the tradeoffs which may be involved in this process. One particular aspect which has featured in the discussion about HEMS since its inception is flight safety. Less than one year after its introduction as the first civilian emergency medical helicopter in Germany, *Christoph 1* crashed on 17 August 1971 due to obstacle strike by the tail rotor on approach of the landing site, killing one occupant and severely injuring two (“Unglücke der Luftrettung”, 2010). More recently, in an accident in the Apennine Mountains in Italy on 24 January 2017 which exemplifies several typical features of fatal HEMS crashes (see below), an air rescue helicopter collided with the mountainside, instantly killing all six occupants of the aircraft (Aerossurance, 2018).

Case reports such as these are further supported by accident statistics relating the number of accidents or casualties to exposure, i.e. hours of flight or number of missions. On average, 5.6 accidents per 100,000 flight hours occurred between 1998 and 2007 in the US, with an average yearly fatality rate of 113 per 100,000 crewmembers. In Germany, the accident rate was found to be 4.6 per 100,000 missions, and a corresponding estimate for the United Kingdom is 5.0 accidents per 100,000 missions (all rates with respect to the comparison period 1998 – 2007, as calculated by the author based on data from Blumen, 2009, Hinkelbein et al., 2011, and Chesters, Grieve, & Hodgetts, 2014). These are rather high rates relative to other types of commercial air transport; in fact, at least in the US in 2007, HEMS crewmembers had among the highest fatality rates in a comparison of several high-risk occupations, including for example police officers, power-line installers, aircraft pilots in general, and logging workers (Blumen, 2009).

Several factors contribute to an increased accident risk of HEMS flights. Unlike in most other types of aviation, operations may need to be conducted in congested, uneven, featureless, or otherwise difficult terrain with an increased risk of terrain collision or obstacle strike (Blumen & UCAN Safety Committee, 2002; Rodenberg, Blumen, & Thomas, 2014). The pilots are also exposed to a unique array of physical and psychosocial stressors during their work, including, for example, vibration, postural and thermal stress; long work hours and shift work; emotionally demanding patient encounters; the necessity to conduct a complex, high-stakes task under time pressure; or the need for quick decision-making in situations where relevant information may be lacking (Carchietti, Valent, Cecchi, & Rammer, 2011; Hickman & Mehrer, 2001; Radstaak, Geurts, Beckers, Brosschot, & Kompier, 2014). Not least due to their inherent urgency, HEMS missions may also carry a higher risk of unplanned adverse weather encounters (Blumen & UCAN Safety Committee, 2002; Butler, 2014; Connell & Reynard, 1995), one of the most prominent causes of fatal HEMS accidents (Baker et al., 2006; Blumen & UCAN Safety Committee, 2002).

In the aforementioned accident in Italy, for example, the investigation report concluded that the pilot may have perceived a sense of urgency because of delays in the loading of the patient and because several prior missions had to be canceled due to bad weather. This may have contributed his continuing the flight according to Visual Flight Rules (VFR) despite poor visibility resulting from fog and the featureless, snowy surroundings, which apparently led to disorientation and ultimately, collision in the mountainous terrain. The pilot’s decision to remain in VFR flight might also have been influenced by a sense of (over)confidence since he was very familiar with the local area and by the fact that he had had only limited recent instrument flight practice (despite formally fulfilling instrument flight training requirements) (Aerossurance, 2018; Agenzia Nazionale per la Sicurezza del Volo, 2018).

In general, in the literature on aviation accident contributory factors, a consistent finding has been that a large majority (from 60 to over 80%) of accidents involve pilot errors and mistakes (e.g., in decision-making, memory, or sensorimotor performance), whereas machine failure nowadays is rather rare (Bledsoe & Smith, 2004; Dambier & Hinkelbein, 2006; Martinussen & Hunter, 2010). However, according to current human factors theories which take a systemic perspective of accident causation, pilot error should be mainly regarded as a proximal cause that can be, and often is, influenced by more distal contextual factors such as inadequate oversight of aircraft operators or stressful working conditions (Reason, 2000; von Thaden, Wiegmann, & Shappell, 2006). Again with respect to the example case from Italy, given that the base was located in a mountainous area with increased risk of inadvertent entry into instrument meteorological conditions (IMC), the accident might have been prevented had the pilot been given additional instrument flight training by the operator, and/or had the base been equipped with one of the several terrain awareness and warning system (TAWS)-fitted helicopters available to the operator at the time (Aerossurance, 2018; Agenzia Nazionale per la Sicurezza del Volo, 2018).

1.3 The role of pilot age and the “Age 60 Rule”

While contextual and organizational factors contributing to aircraft accidents have received attention only more recently, the significance of pilot age for flight safety has been a controversial issue in the aviation sector for decades (Aerospace Medical Association, 2004). On the population level, there is a clear association of older age with declines in cognitive (Salthouse, 2004), psychomotor (Era et al., 2011), and sensory (Swenor, Ramulu, Willis, Friedman, & Lin, 2013) function, as well as an increase in the risk of onset of medical conditions in general (Barnett et al., 2012) and also specifically of conditions with a potential for sudden incapacitation, such as myocardial infarction, stroke, or epilepsy (Hauser, Annegers, & Kurland, 1993; Mozaffarian et al., 2015). These well-known facts seem to give plausible reason for concern about older-age professional pilots, considering the high job demands placed on them and the potentially disastrous consequences of their failure to perform adequately or of in-cockpit sudden incapacitation. Attention to the issue is further magnified by the ongoing population aging in many countries across the globe (United Nations Department of Economic and Social Affairs Population Division, 2015).

The aforementioned concerns were cited in justification of the “Age 60 Rule” imposed by the US Federal Aviation Administration (FAA) in 1959, which forbade pilots aged 60 and over to engage in air carrier operations (Federal Aviation Administration, 1959). Variants of this original Age 60 Rule have since been introduced, and at times modified, by different regulatory authorities. For example, the International Civil Aviation Organization (ICAO) of the United Nations introduced an age limit of 60 for the pilot-in-command (PIC) in international commercial air transport as a recommendation in 1963, and changed this to a mandatory standard in 1972. In 2006, the ICAO rule was relaxed to allow one pilot to be up to 64 years old in a two-pilot setting where the co-pilot is less than 60 years old (Evans, 2011), and since 2014, both pilots of a multi-pilot flight may be up to 64; however, the age limit for single-pilot operations is still 59.

Change histories such as these attest to the controversial nature of the Age 60 Rule(s), which opponents have criticized as ageist and politically motivated (DuBois, 2005; Wilkening, 2002). Instead of a general age limit, flight safety could be protected by additional or more refined medical, psychological, and job performance testing procedures to identify pilots who are affected by age-related pathology and performance decline (Stuck, van Gorp, Josephson, Morgenstern, & Beck, 1992). This line of reasoning is (implicitly or explicitly) based on the notion of “biological age” (as opposed to chronological age) which is assumed to differ between individuals of the same age – what is colloquially referred to as “aging poorly/well”. Indeed, this concept is arguably a cornerstone of gerontology, although a notoriously elusive one, and many attempts have been and are being made in this discipline to measure it (Fried et al., 2001; Lowsky, Olshansky, Bhattacharya, & Goldman, 2014; Rowe & Kahn, 1997).

Although the Age 60 Rule's opponents' argument is at least implicitly based on the concept of biological age(ing), it has apparently not been directly applied in the field of aerospace medicine so far. However, professional pilots do have to pass a strict selection and training procedure to obtain initial licensing, and undergo regular proficiency and medical fitness checks to maintain the license (Evans, Evans, & Harper, 2016; Martinussen, 2017). This not only suggests that they are healthier than the general population, which is confirmed by their considerably lower morbidity and mortality rates with respect to a whole range of illnesses (Hammer et al., 2014; Linnersjö, Brodin, Andersson, Alfredsson, & Hammar, 2011). It may in fact be the case that older age pilots have an especially good health status and performance record since they managed to pass recurrent testing despite their age – a process known as "healthy worker survivor effect" (Arrighi & Hertz-Picciotto, 1994). Moreover, older pilots' greater on-the-job experience and the associated increase in implicit and explicit domain-specific knowledge and skills may further mitigate or even outweigh any age-related sensorimotor or cognitive performance declines (Salthouse, 2012). This likely explains why no clear association between pilot age and performance was found in literature reviews (Aerospace Medical Association, 2004; Hardy & Parasuraman, 1997).

Finally, the age limit of 60 years is rather arbitrarily chosen; for example, in the general population, cardiometabolic risk indicators (Hardy, Lawlor, & Kuh, 2015) and cardiovascular mortality (Mikkola, Gissler, Merikukka, Tuomikoski, & Ylikorkala, 2013; Vaidya, Becker, Bittner, Mathias, & Ouyang, 2011) start to increase notably well before the age of 60, and no threshold effect around age 60 is visible which would empirically justify this limit as opposed to, say, 55 or 65 years.

Despite these arguments, the Age 60 Rule was made European Union law through European Union Regulation 1178/2011, point FCL.065 of Annex I, which required at least one pilot in commercial air transport operations to be less than 60 years old. While this mimicked the corresponding requirement of the now-defunct Joint Aviation Authorities (JAA; predecessor organization of the European Aviation Safety Agency, EASA), JAA regulations were binding only for cross-country flights, thereby allowing for some flexibility regarding pilots' age within national airspaces. For example, according to §127 LuftPersV (Verordnung über Luftfahrtpersonal), holders of a German license were allowed to fly single-pilot operations until the age of 65 years, as long as they remained within German territory. Based on the rationale to establish a harmonized regulatory framework across European countries, such national exceptions were invalidated by EU Regulation 1178/2011.

Although the legislative change hardly affected aviation operations where multi-pilot crews are routinely employed (most notably airline transport), it poses a significant problem to European HEMS operators since HEMS flights are often operated by a single pilot due to space and takeoff weight limitations. HEMS operators have therefore expressed concern over the resulting loss of highly experienced personnel (Poguntke, 2012). Recruitment and training of new HEMS pilots is a lengthy process due to demanding entry requirements, and currently several operators rely on temporary exemptions from the Age 60 Rule for their affected pilots. Besides the potential adverse effect of EU Regulation 1178/2011 on HEMS operations, it may – depending on the retirement rules of the respective country – also put aging HEMS pilots into a precarious personal situation since grounding due to the rule is often not covered by loss-of-license insurances, and since only limited alternative employment opportunities are available (e.g., working as a flight instructor).

1.4 Overview of the "Age 60" project

1.4.1 Study objectives

In response to EU Regulation 1178/2011, a German HEMS operator contacted the Institute and Clinic for Occupational, Social, and Environmental Medicine at the University Hospital of LMU Munich with a request to evaluate the validity of the scientific rationale of FCL.065. This led to an initial "Age 60" study on the relation of HEMS pilots' age to their medical and cognitive fitness which was conducted by researchers of the institute in cooperation with one Austrian and two German HEMS operators in 2012/13.

To address limitations of this initial study raised by EASA (including small sample sizes, particularly for pilots aged 60 and over, lack of medical data of HEMS pilots, and uncertain generalizability of results stemming only from German and Austrian pilots to the European context), an extension study, funded by the European HEMS & Air Ambulance Committee e.V. (EHAC) was carried out in

2015/16 in cooperation with the three HEMS operators who had already participated in the initial study plus one Czech and one Polish operator. The data collected in the course of this extended Age 60 Study form the basis of the present thesis. The main objective of the study was to determine if and how the current age limit according to FCL.065 can be raised without a significant increase in the associated risk.

1.4.2 Study design and results

The initial Age 60 Study was designed according to a multi-method approach to investigate the relation between pilot age and cognitive/medical fitness from different complementary perspectives. Design and results of these different study tiers can be summarized very briefly as follows:

1. A systematic review of literature on professional pilot age and incapacitation found that in-flight incapacitation is a very rare event (occurring at a rate of 0.19–0.45 per 10^6 flight hours), and that incapacitation does increase with age; however, only one in-flight incapacitation study actually included pilots aged 60 and over, and found no such event in this group; also, no data on HEMS pilots were available (Huster, Müller, Prohn, Nowak, & Herbig, 2014)
2. According to a quasi-experimental investigation of HEMS pilot performance in managing two malfunction scenarios enacted in a helicopter flight simulator, performance showed either no or a curvilinear U-shaped relationship with age, such that younger and older pilots performed better than middle-aged pilots
3. Ratings by flight examiners from actual check flights (“line checks”) regarding several aspects of a pilot’s flight competence (e.g., proper pre-flight briefing, working through check lists and procedures, communication behavior during the flight, execution of flight patterns) available for 91 pilots did not reveal any association with age
4. An analysis of 1,770 liability damage cases resulting from operations carried out by 257 pilots of the three operators between 2007 and 2011 found a small ageing-related (=longitudinal) effect such that the number of damage cases increased over time in younger, but not in older pilots (Müller et al., 2014)

Thus, on the whole, there was little evidence from the initial Age 60 study that safety is compromised in HEMS operations with older pilots, although as mentioned in the previous section some study limitations remained. The extension study aimed to complement the findings of the initial study, and thereby address its limitations, by collecting data from three distinct sources:

1. Additional HEMS pilot flight simulator performance ratings collected during pilots' training/check flight sessions, replicating the design of the initial study for data pooling; in the course of the simulator session, information on pilots' experience of the session was collected by a short questionnaire as in the initial study; a newly added longer questionnaire on the pilots' working conditions in general inquired about work-related physical and psychological stressors, as well as physical and psychological symptoms of strain
2. Pilots' aeromedical examination records covering the previous 10 years, requested from aeromedical examiners (AMEs) and/or aeromedical centers (AeMCs); professional pilots need to undergo such examinations every 6 months if they are either 60 years and over or 40 years and over and conduct single-pilot operations (every 12 months otherwise).
3. Anonymized longitudinal staff records from the participating operators, including information on the employed pilots' age and their involvement in liability damages; for pooling with the corresponding data from the initial study

HEMS pilots of the five operators with an upcoming simulator flight session during the study data collection period (September 2015 to October 2016) at one of two training sites in Hangelar, Germany, and Warsaw, Poland, were identified by the operators, and were asked for participation either directly during the training session by a research team member or a flight instructor, or prior to the session via the employer's internal communication channels. In the recruitment of participants, priority was given to older-age pilots (55 and over). Additionally, a small number of pilots who did not attend a training session at the two sites during the data collection period were asked for participation (medical study part and working conditions questionnaire only). The pilots could separately indicate their consent to participate in the flight simulator and/or the medical study part. In the latter case, the pilots were additionally asked to release their AMEs/AeMCs from their duty of non-disclosure. Signed release forms with a request for transfer of aeromedical examination documents were then mailed to all specified physicians/centers.

Simulator performance, questionnaire, and aeromedical records data were collected in personal, i.e. non-anonymous, form to allow data linkage. This procedure had been explicitly prespecified in the study protocol, which had been approved by the Ethics Committee at LMU's Faculty of Medicine (Project No. 466-15) prior to data collection. The participant flow for the study parts involving personal data is shown in Figure 1. All participants were male; this was because the operators' pilot workforces consisted virtually exclusively of men at the time of data collection. The medical data, obtained from 24 distinct sources, included records from a total of 977 aeromedical examinations conducted between 2004 and 2016 (133 from pilots aged 60 and over at the time of their last available examination), with considerable variation between pilots in follow-up time (average 8.52 years, range: 0-12, median: 9.59) and between sources in the volume of documents per examination.

Anonymized liability damages records were collected directly from the HEMS operators without any involvement of pilots. Upon pooling with the data available from the initial Age 60 Study, data from 353 pilots (22 of which were 60 and over at the end of follow-up) covering a total of 1,592 pilot-years between 2007 and 2015, during which 1,853 incidents were recorded, were available for an updated liability damages analysis.

A comprehensive study report was submitted to EHAC in December of 2016 and then forwarded further to EASA. The results informed a continuing stakeholder dialogue on possible reforms of FCL.065 and other regulatory provisions pertinent to the pilot age issue, e.g. regarding operational safeguards. For example, the findings were presented and discussed at the 10th EASA Rotorcraft Symposium (December 2016) and at a March 2017 workshop on the Age 60 Rule organized by the Austrian civil aviation authority; both events included representatives of HEMS operators and regulatory agencies as well as researchers from different European countries. In the remainder of this section, study results which are scheduled to be published in scientific journals, but which are not submitted as part of the thesis are further described. The publications included in the present thesis are described in the next section.

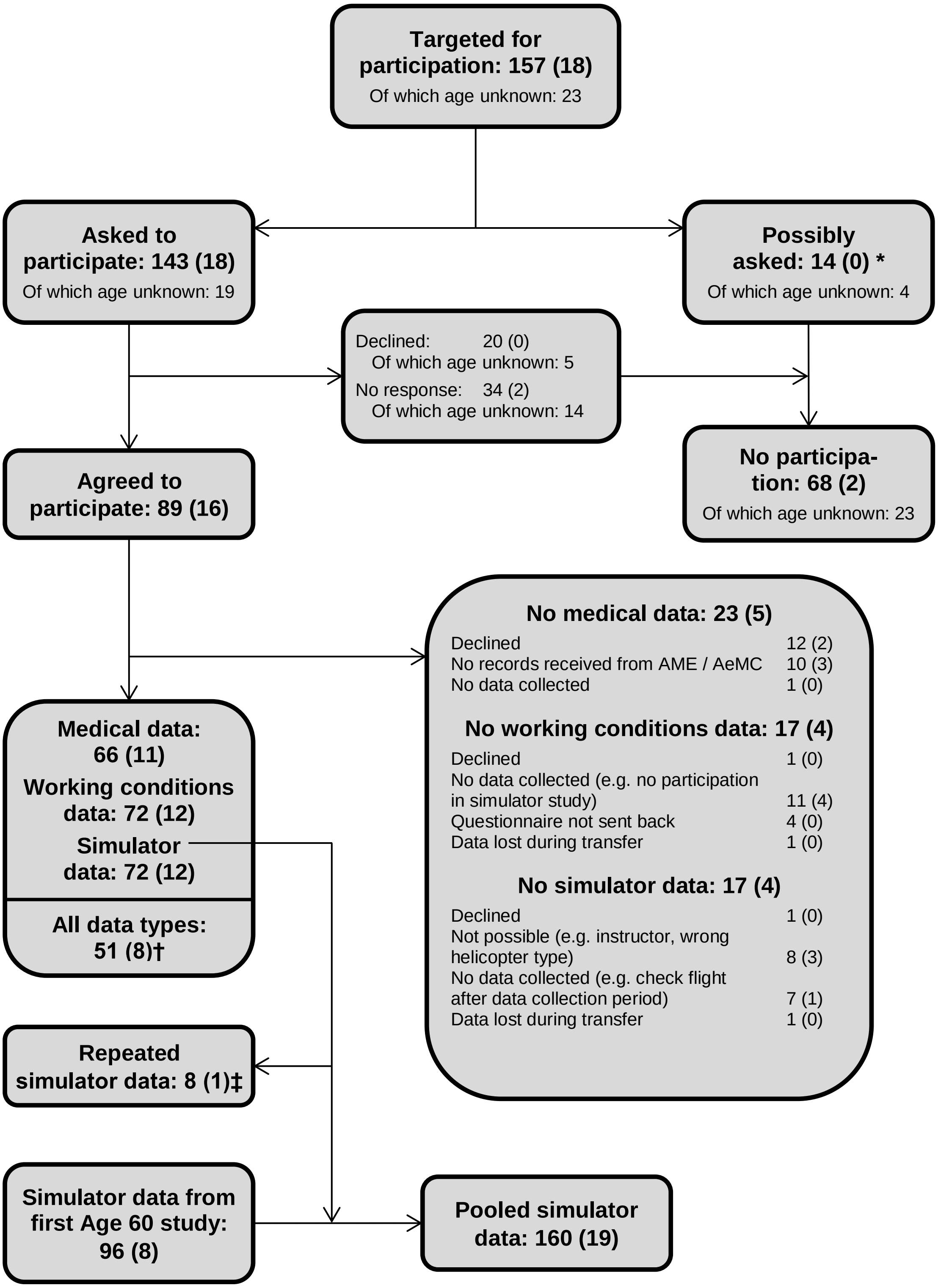


Figure 1. Participant flow. Number of participants aged ≥60 in parentheses (for extended Age 60 Study, based on age as of 31 December 2015 – actual age during data collection may differ somewhat). *Unknown if request for participation by HEMS operator / instructor was actually made. †Participants where all 3 types of data are available. ‡Simulator data collected both in initial and in extended Age 60 Study.

Analyses of simulator performance ratings for the pooled sample of 160 German, Austrian, Polish, and Czech HEMS pilots (manuscript currently in preparation) broadly confirmed the findings of the initial Age 60 Study: Either no or inverse U-shaped effects of age were apparent for the two study scenarios. Overall, the results suggest that the required sensory, psychomotor, and cognitive capabilities of HEMS pilots near the critical age threshold of 60 years do not appear to be affected in a relevant degree by their age and in fact may be superior to those of some of their middle-aged colleagues. Possible explanations for these findings are their greater domain task experience (Aherne, Zhang, & Newman, 2016) and a healthy worker survivor effect (Arrighi & Hertz-Picciotto, 1994) due to the required recurrent medical and performance testing to maintain the pilot license. Moreover, similar to the simulator data, the analysis of the pooled liability damages data also confirmed the results of the initial Age 60 Study: although overall incidents increased over time, this increase tends to be lower in older pilots.

Further evidence for a healthy worker survivor effect was obtained in the medical study part by an analysis of the association of pilot age with an index of physiological dysregulation constructed from 18 biomarkers (submitted to *International Archives of Occupational and Environmental Health*; currently under review). The rationale for this analysis is grounded in recent biogerontological research which found that (a) indices of biological age(ing) or physiological dysregulation derived from multiple biomarkers predict mortality, morbidity, and physical and cognitive impairment (Arbeev et al., 2018; Belsky et al., 2015; Seeman, McEwen, Rowe, & Singer, 2001), and (b) the exact choice of biomarkers used to construct such indices is not decisive, provided that a sufficient number of markers and organ systems is sampled (Cohen et al., 2015; Rockwood & Mitnitski, 2007; Seplaki, Goldman, Glei, & Weinstein, 2005). Thus, the 18 biomarkers used were chosen based on their association with morbidity and mortality (as determined by literature review), their availability in the received aeromedical examination records, and such that a broad range of organ systems were represented (e.g., cardiovascular system: systolic blood pressure; metabolic system: fasting plasma glucose; immune system: white blood cell count). Although longitudinally, an increase in physiological dysregulation scores was observable over an average follow-up time of 7.9 years, cross-sectionally there was an inverse U-shaped relationship between age and dysregulation in 52 HEMS pilots, such that dysregulation was estimated to be highest around 45 to 50 years. Assuming that our dysregulation index captures risk of onset of overt pathology (Milot et al., 2014), its cross-sectional decrease after age 50 again suggests a selection process such that pilots where pathology begins to manifest do not pass medical recertification or leave the profession voluntarily. Overall, the findings from the simulation and the dysregulation analyses therefore suggest that the rationale of the Age 60 Rule may be flawed as it does not consider such selection effects.

As discussed previously, HEMS pilots' work is demanding and involves a potentially large number of stressors at work, such as time pressure, decision-making under uncertainty, and witnessing human suffering. Especially in combination with organizational stressors like lack of management support or contradictory work demands, these conditions can lead to psychological strain responses such as reduced well-being, job satisfaction and motivation, or even manifest psychological disorder (Humphrey, Nahrgang, & Morgeson, 2007; Stansfeld & Candy, 2006). Such strain responses are in turn negatively associated with performance and safety at work (Judge, Thoresen, Bono, & Patton, 2001; Nahrgang, Morgeson, & Hofmann, 2011). The prevalence of work stressors in HEMS pilots, as well as their relation to psychological strain, is therefore an important but so far hardly investigated aspect of flight safety in HEMS. In an analysis of the responses to the working conditions questionnaire of 72 HEMS pilots from the extended Age 60 Study (submitted to *Air Medical Journal*; currently under revision), we found that overall the pilots reported a quite favorable profile of work stressors (e.g., role conflict or work pace) and resources (e.g., social support, procedural justice), both when compared to the general working population and to airline pilots (based on data taken from literature). However, in those pilots who reported the presence of work-family conflict or perceived procedural injustice, the encounter of emotionally disturbing situations, and a lack of role clarity at work, notably reduced levels of well-being and energy were observable. The results may be helpful in the implementation of measures to promote pilot mental health (Aerospace Medical Association, 2016) in the context of HEMS; for example, the great pride and dedication which the HEMS pilots invest into their occupation (as a resource, but also as a precondition for perceived organizational injustices) should be taken into account as well as the need to discuss emotionally burdensome events such as failed rescue missions.

1.5 Thesis publications

1.5.1 Thesis objectives

The present thesis is inextricably linked to the extended Age 60 Study project, and therefore one of the thesis objectives was to evaluate the significance of the 60 years threshold, as stipulated in FCL.065, to flight safety in HEMS. However, the focus of the thesis is not only on regulatory issues but also more generally on evaluating the significance of different aspects of flight safety in HEMS. Finally, the thesis also aimed to apply novel methods in the field of aviation safety to explore possibilities of expanding the toolbox available to researchers in the field.

1.5.2 Summary of original publications included in this thesis

One major focus of concern regarding medical fitness of older-age pilots which has repeatedly been emphasized by Age 60 Rule proponents (also in the recent discussion around FCL.065) is the risk of sudden in-cockpit incapacitation due to major cardiovascular events such as myocardial infarction or stroke. The increase in general population cardiovascular mortality and morbidity rates associated with age has been cited in support of this argument. However, as previously discussed, the cardiovascular risk of professional pilots is considerably lower than that of the general population; age-specific figures for absolute risk of cardiovascular events in professional pilots' are rare, and absent for HEMS pilots specifically. In Publication 1, which is based on the aeromedical records data, we therefore longitudinally investigated age-related change of six markers of cardiometabolic risk (systolic blood pressure, electrocardiogram QTc interval, body mass index, total serum cholesterol, high-density lipoprotein cholesterol, fasting blood glucose), and of fatal cardiovascular event risk as estimated by an established scoring algorithm (SCORE), in 66 HEMS pilots over an average follow-up period of 8.5 years, and compared these changes between those pilots aged 60 and over and those less than 60 years at the end of follow-up (termed "older pilots" and "younger pilots", respectively). Overall, the pilots' risk of a fatal cardiovascular event within six months was estimated between 0 and 0.3%, and increased notably with age. However, the relative increase in risk scores over time was faster in the younger, compared with the older pilots, and increases in BMI and fasting glucose observed during follow-up in the younger pilots significantly decelerated in the older pilots; also, the lipid profile improved over time in the older pilots, whereas it did not change appreciably in the younger pilots. Although two markers of cardiac risk tended to increase more strongly in the older pilots, these differences did not reach statistical significance. Finally, large proportions of the variation in risk markers were attributable to unobserved time-stable interindividual differences, rather than to the observed variables. In summary, the data do not give much empirical support to a general age threshold of 60 years regarding cardiovascular fitness of HEMS pilot. A short-term solution to balance staffing needs of HEMS operators (and therefore adequate HEMS coverage) and safeguards against cardiovascular risks might be to conduct a coronary artery calcium (CAC) scan at age 60 to guide the decision whether a pilot may continue flying single-pilot, as this method captures cumulative effects of degenerative and pathological vascular processes and has a very high negative predictive value. In the long term, individualized risk scoring using an appropriately calibrated prediction algorithm should be systematically applied well before age 60.

In Publication 2, a wider perspective is taken with respect to factors affecting flight safety in HEMS. As illustrated in section 1.2 with the example case from Italy, accidents in aviation are best understood as resulting from an unfavorable constellation of circumstances both at the level of the pilot (e.g., fatigue, medical and psychological fitness), proximal circumstances (e.g., weather conditions, workplace stress), and the wider organizational and regulatory environment (e.g., inadequate oversight). This complexity and the resulting wealth of potential data sources which may be utilized to identify hazard and protective factors calls for an application of novel analytic methods that can handle large numbers of predictors (even when sample size is relatively small, as is likely the case with small target populations such as HEMS pilots), and that do not impose restrictive model assumptions, such as linearity of effects. In particular, data mining and machine learning approaches, which have become popular recently mainly due to increases in available computational power, may be leveraged in the field of aviation safety. We explored the utility of Random Forests, a supervised machine learning method based on decision trees, by selecting the most powerful predictors of simulator flight performance in 51 HEMS pilots where data from all sources – performance ratings, aeromedical data and questionnaires – were available. The overall set of candidate predictors consisted of 54 variables including medical risk markers, physical and psychosocial work stress and strain indicators, pilots' subjective experience of the simulator session, and general pilot and simulator characteristics. Five predictors were selected by the method as predicting performance above chance level: Perceived rewards at work, perceived predictability of work demands, physiological dysregulation, alanine aminotransferase, and simulator type. The direction of effects was theoretically plausible; for example, higher perceived rewards were associated with higher performance, whereas higher dysregulation was associated with lower performance. Although these results are of an explorative nature, they demonstrate the potential of the Random Forest method for aviation safety studies. For example, the method could be applied to information from accident investigation databases to identify factors associated with accident lethality, or to analyze effects of operational conditions (e.g., timing of missions, weather, geographical location) on mission safety parameters based on HEMS providers' administrative data.

1.5.3 Contribution of thesis author to included publications

For both included publications, I had primary responsibility for data collection, data management & quality control, data analysis, and drafting of the manuscript. I also assisted in the formulation of the extended Age 60 Study's overall design, as outlined previously.

2. Publication 1: Age(ing) and cardiometabolic risk in HEMS

Bauer, H., Nowak, D., & Herbig, B. (2018). Aging and cardiometabolic risk in European HEMS pilots: An assessment of occupational old-age limits as a regulatory risk management strategy. *Risk Analysis*, *38*, 1332–1347. https://doi.org/10.1111/risa.12951

Aging and cardiometabolic risk in European HEMS pilots: an assessment of occupational old-age limits as a regulatory risk management strategy

Hans Bauer[a,]*, Dennis Nowak[a], & Britta Herbig[a]

[a]Institute and Clinic for Occupational, Social, and Environmental Medicine, University Hospital of LMU Munich, Germany

*Corresponding author. Postal address: Institute and Clinic for Occupational, Social, and Environmental Medicine, Ziemssenstraße 1, 80339 München, GERMANY. Phone: +49 89 4400 55312, Fax: +49 89 4400 54564, Email: hans.bauer@med.uni-muenchen.de

Funding & conflict of interest statement

This paper reports findings of a study on age and flight safety in helicopter emergency medical services (HEMS) pilots funded via an unrestricted educational grant from the European HEMS & Air Ambulance Committee e.V. (EHAC). EHAC represents European air rescue organizations who favor a liberalization of Age 60 regulations. Freedom from interference in any stage of the research process was laid down contractually between EHAC and the research project team, represented by the manuscript authors. EHAC had no role in the formulation of the study design, the collection, analysis and interpretation of the data, the writing of the report, and the decision to submit the paper for publication.

Acknowledgements

We thank all participating HEMS operators and their representatives for their support of the study. We also thank EHAC for providing background information on the subject. Moreover, we gratefully acknowledge the efforts taken by the participating HEMS pilots and the trust they placed in us.

ABSTRACT

Background

Old-age limits are imposed in some occupations in an effort to ensure public safety. In aviation, the “Age 60 Rule” limits permissible flight operations conducted by pilots aged 60 and over. Using a retrospective cohort design, we assessed this rule’s validity by comparing age-related change rates of cardiometabolic incapacitation risk markers in European helicopter emergency medical service (HEMS) pilots near age 60 with those in younger pilots.

Methods

Individual clinical, laboratory, and electrocardiogram (ECG)-based risk markers and an overall cardiovascular event risk score were determined from aeromedical examination records of 66 German, Austrian, Polish, and Czech HEMS pilots (average follow-up 8.52 years). Risk marker change rates were assessed using linear mixed models and generalized additive models.

Results

Body mass index increases over time were slower in pilots near age 60 compared to younger pilots, and fasting glucose levels increased only in the latter. Whereas the lipid profile remained unchanged in the latter, it improved in the former. An ECG-based arrhythmia risk marker increased in younger pilots, which persisted in the older pilots. Six-month risk of a fatal cardiovascular event (in or out of cockpit) was estimated between 0 and 0.3%. Between 41 and 95% of risk marker variability was due to unexplained time-stable between-person differences.

Conclusions

The cardiometabolic risk marker profile of HEMS pilots appears to improve over time in pilots near age 60, compared to younger pilots. Given large stable interindividual differences, we recommend individualized risk assessment of HEMS pilots near age 60 instead of general grounding.

KEYWORDS

Cardiovascular risk; Flight safety; Government regulation

1. INTRODUCTION

1.1. Background

1.1.1. Occupational old-age limits as a regulatory risk management strategy

Workforce aging associated with the general global trend towards older populations(1) has led to an intensified focus on the capabilities, performance, and safety of older workers, particularly in those whose occupational activities directly involve the safety and well-being of third parties (e.g., passengers, patients, or the community).(2–4) A customary regulatory approach towards limiting public safety risks associated with a potential age-related performance decline has been the imposition of early mandatory retirement ages for certain occupations,[1] such as professional pilots, air traffic controllers, policemen, firefighters, public officials, or technical supervisory personnel.(5,11,12) For example, in the United States, retirement is mandated at 57 years for federal law enforcement officers, firefighters, or nuclear materials couriers, and at 56 years for air traffic controllers (5 U.S. Code § 8335). In Germany, retirement ages for federal police officers and air traffic controllers are 62 (BPolBG § 5) and 55 (based on a collective labor contract), respectively.

Aviation is an occupational field where the topic of mandatory retirement has featured prominently since decades, as evidenced by the history of the "Age 60 Rule".(13,14) First introduced in 1959 by the United States Federal Aviation Administration, it prohibited air carrier operations by pilots aged 60 and over, and was subsequently adopted (in modified form) in other countries and also by the International Civil Aviation Organization (ICAO). Ever since its adoption, the rule has been the cause of controversy, being criticized as ageist and inequitable.(7) Given continuing improvements in health status and cognitive capacities of successive generations of older adults,(15) the wide variability in health and functioning between individuals of a given cohort,(16,17) and absent or inconsistent associations between pilot age and actual flight performance,(13) it has frequently been argued that

---

[1] It should be noted that political and economic considerations, such as the defense of advantageous early retirement schemes, labor market access of the younger generation, or company profitability, have also played a prominent role in the establishment and maintenance of mandatory retirement rules both in the public and private sector.(5) Thus, both between and within involved entities such as authorities, industry, and employees, one may find subgroups opposing or supporting age-based mandatory retirement.(6,7) The ambivalence of this issue is mirrored in the larger-scope societal disputes about the requirement (privilege) to continue working (retire) at old ages(8–10)

general age limits for professional pilots should be replaced by fitness- and performance-based testing focused on risks related to aging.(18,19)

Although general age limits for professional pilots have been relaxed in recent years in response to such criticism, recent regulatory decisions illustrate that authorities still consider such limits to be necessary. For example, since 2014, ICAO standards permit multi-pilot international commercial air transport flights with pilots up to age 65 (Amendment 172 to Annex 1 of the Chicago Convention), and the same holds within the European Union (Commission Regulation 1178/2011, as amended by Commission Regulation 2015/445; since 2015). However, in both cases, the age limit of 60 is upheld for single-pilot operations. Consistent with this prescription, an analysis of the legal situation in Germany and Norway concluded that single-pilot operations constitute a case where mandatory retirement provisions may justifiably prevail over existing anti-discrimination law.(5)

One concern which is raised frequently in this context is that medical-cause sudden incapacitation occurring in a solo pilot during flight cannot be compensated by a second pilot, thereby amplifying the risk potential of age-related health deficits, in particular cardiovascular events such as myocardial infarction or stroke.(6,20,21) Although the argument seems plausible at first glance due to the well-established age-related increase of cardiometabolic risk in the general population,(22) the choice of the age threshold is essentially arbitrary – at the population level, cardiometabolic risk markers(23) and cardiovascular mortality(24) start to increase well before age 60; for example, the 2015 Global Burden of Disease study estimated the annual cardiovascular mortality in the United States at 71, 124, 194, and 287 per 100,000 persons at ages 45–49, 50–54, 55–59, and 60–64, respectively.(24) Furthermore, applicability of findings in the general population might be restricted for professional pilots, who are tightly monitored and display a markedly lower cardiovascular morbidity and mortality, with age-standardized mortality ratios relative to the general population usually at or below 0.5.(25,26) The only available systematic review on age and medical-cause incapacitation in professional pilots indicated an age-related increase; however, only one of the reviewed studies included active pilots aged 60 and over, and found no adverse events in this group.(27)

1.1.2. European helicopter emergency medical services and the "Age 60 Rule"

Prior to European Union Commission Regulation 1178/2011, the Age 60 Rule for single-pilot operations was legally binding only for cross-border flights within Europe, but now applies to all commercial air transport flights. This has especially affected European helicopter emergency medical services (HEMS) operators, since HEMS flights are often operated by a single pilot due to space and takeoff weight limitations. The regulation has therefore been criticized by HEMS operators, who have expressed concern over the loss of highly experienced personnel and the endangerment of comprehensive HEMS coverage in Europe.(28) Recruitment and training of new HEMS pilots is a lengthy process due to demanding entry requirements, and currently several operators rely on temporary exemptions from the Age 60 Rule for their affected pilots. Besides the potential adverse effect of the Age 60 Rule on HEMS operations, it may – depending on the retirement rules of the respective country – also put aging HEMS pilots into a precarious personal situation since grounding due to the rule is often not covered by loss-of-license insurances, with limited alternative employment opportunities (e.g., working as a flight instructor).

Although HEMS pilots are held to the same standards of medical fitness as other professional pilots, they are subjected to a unique array of occupational demands and stressors including vibration, noise, thermal and postural stress, shift work, long working hours, unpredictable mission schedules, and emotionally demanding emergency patient encounters.(29–33) They are required to conduct a complex high-stakes task under time pressure and potentially adverse environmental conditions.(34) The accident rate in HEMS flights is also relatively high compared to other types of commercial air transport, with an average accident rate of 5.6 per 100,000 flight hours and an average yearly fatality rate of 113 per 100,000 crewmembers between 1998 and 2007 in the United States (authors' own calculation based on data from I. Blumen(35)). In Germany, the accident rate in the same time span was found to be 4.6 per 100,000 missions (authors' own calculation based on data from J. Hinkelbein et al.(36)). The degree to which age-related medical-cause sudden incapacitation risk plays a role in these figures is, however, unknown.

1.2. Study objective

In an effort to further evaluate the empirical basis of the Age 60 Rule, we therefore aimed to assess age-related longitudinal change rates of common individual cardiometabolic risk markers, that is, systolic blood pressure (SBP), serum total cholesterol (TC), serum high-density lipoprotein cholesterol (HDLC), fasting glucose, body mass index (BMI), and heart rate-corrected electrocardiogram (ECG) Q-wave – T-wave interval (QTc interval), as well as an established summary score of overall cardiovascular risk, over approximately ten years in European HEMS pilots. Elevated SBP and TC and decreased HDLC are associated with the development of atherosclerotic cardiovascular disease.(37) Elevated fasting glucose and BMI are associated with the development of type 2 diabetes mellitus and thereby indirectly (though likely also directly) with cardiovascular disease.(38,39) A prolonged QTc interval is predictive of cardiac arrhythmia and sudden death.(40) Cardiovascular event occurrence is strongly related to common risk factors such as smoking and blood pressure in professional pilots (as in the general population).(41) Investigating cardiometabolic risk markers therefore offers a feasible way to study the incapacitation risk in the small and healthy population of active HEMS pilots. In the analysis, we focus on comparing the direction and magnitude of risk marker change rates between pilots approaching the critical age threshold of 60 years and the remaining pilots.

2. METHODS

2.1. Study design and setting

In this retrospective cohort study we analyzed mandatory aeromedical examination records of active HEMS pilots employed in 2015/2016 with one of five air rescue organizations (two based in Germany and one each in Austria, Poland, and the Czech Republic) operating about two thirds of all HEMS bases in these countries. Examination records were requested from physicians or medical centers which the consenting pilots had visited for medical certification during the ten-year period prior to study participation.(42) Examinations include personal and family history, tobacco, alcohol and medication use, a clinical inspection, and a number of basic clinical measurements (e.g., anthropometrics, blood pressure, hemoglobin). Intermittently, further examinations such as resting

ECG are required. Further procedures can be requested as deemed necessary by the examiner. This often includes laboratory markers related to cardiovascular risk (e.g., blood lipids), although there is currently no regulatory prescription for a periodic measurement of these markers.

Participant recruitment took place within the overarching framework of a study on age and flight safety which also investigated pilot simulator performance and flight incidents.(43) The study had been approved by the Ethics Committee at the Medical Faculty of LMU Munich (Project No. 466-15), and written informed consent had been obtained from all study participants, prior to data collection.

2.2. Participant selection

During the data collection period from September 2015 to October 2016, active HEMS pilots from the five air rescue organizations completing a check or training flight session at training sites in Hangelar, Germany, or Warsaw, Poland, were either personally approached for participation by a research team member or a flight training instructor of the respective organization, or were contacted via their organization’s internal communication channels. Recruitment was consecutive with an additional effort to oversample pilots near or above age 60 by identifying and contacting these pilots in advance. Pilots who agreed to participate in the medical study part were asked to provide names and addresses of the physicians and medical centers, and release them from their non-disclosure duty concerning aeromedical examination records from the relevant period of time.

2.3. Data collection & processing

Signed release forms with a request for transfer of full aeromedical examination documents were mailed to all specified physicians/centers, and a reminder letter was sent after four weeks. If the reminder also did not elicit a response, we made at least one more contact attempt via telephone or email. We received documents from 23 physicians/centers, dated between April 2004 and July 2016. Documents were partly in electronic and partly in paper form and can be categorized into the following main classes: (a) standardized examination application and report forms according to European regulations, (b) laboratory measurements, (c) printouts of examination procedures (e.g., ECG), and (d) other medical examiner report forms or questionnaires (e.g., from military). They were in Czech, Polish, or German language but were either in a known standardized format or used standard medical terminology.

All documents were searched for information concerning the variables of interest (see next section). Data entry was conducted according to a coding manual developed after a review of the received documents prior to data entry and continuously revised during the data entry process. Both manual and algorithmic quality control of entered data (plausibility and consistency checks) was conducted.

2.4. Analysis variables

2.4.1. Exposure variables

We calculated age at examination as time from date of birth to examination date. For this purpose, we grouped all documents into half-year periods (“examinations”) from January to June and July to December, based on the actual dates of the corresponding examination procedures, and then assigned to each document the average date of the set of all documents within its respective half-year (“examination date”). This allowed the matching of examination procedures in close temporal proximity while maintaining a reasonable degree of temporal resolution. Time since first available assessment was determined as the difference between age at current examination and age at first available outcome assessment. We further classified each pilot as either “older” (aged 60 and over at last available examination) or “younger” (all remaining).

2.4.2. Confounding and auxiliary variables

Based on the pilot’s country of birth, we defined region of Europe as “East Europe” (Czech Republic, Poland) or “West Europe” (Austria, Germany). Smoking status was defined as a binary variable where the pilot was classified as “current smoker” if in any document of a given examination, any indication of smoking was made. We logically imputed smoking status for examinations without smoking status information by assigning the value of the temporally closest examination with available information. Current use of blood pressure- and of lipid-lowering drugs were defined as binary variables, and coded and imputed in an analogous manner as described for smoking status.

2.4.3. Outcome variables

Resting SBP and BMI information was obtained from physician report forms. BMI was re-calculated from height and weight information whenever possible. Serum TC and HDLC as well as fasting plasma glucose measurements were abstracted from laboratory reports. QTc interval was calculated from uncorrected QT interval and RR interval (=inverse heart rate) information on ECG printouts wherever possible, using Bazett's formula ($QTc(ms) = \frac{QT(ms)}{\sqrt{RR(s)}}$); otherwise, QTc values as provided on the documents were used. If several measurements of a biomarker had been taken over the six-month period of a given examination, we used their arithmetic mean. Fatal cardiovascular event risk within the next six months was estimated using the SCORE (Systematic Coronary Risk Evaluation) equation(44) based on sex, region of Europe, age, smoking status, systolic blood pressure, and total cholesterol. Using a six-month period instead of the customary ten-year period gives a more accurate picture of the absolute risk of a fatal cardiovascular event in between HEMS pilots' aeromedical examinations.

2.5. Statistical Analysis

2.5.1. Descriptive analyses

We calculated descriptive statistics for the first and last available assessment of each outcome variable as well as the average follow-up time between these assessments (follow-up time varied both between outcomes and pilots), stratified by region of Europe and age category.

2.5.2. Main analyses

For each outcome variable, we fitted a linear mixed model (LMM)(45) including a random intercept and fixed effects for region of Europe, age at first available outcome assessment, and time since first available assessment, to account for age and aging effects. SCORE risk was log-transformed prior to analysis to normalize its strongly skewed distribution, but results are presented in the natural scale (thereby assuming multiplicative covariate effects). Models of SBP and cholesterol additionally included, respectively, current blood pressure-lowering and lipid-lowering drug use as covariates. The SCORE risk model included a variable indicating current use of either of these two drug types. To test if any differences exist in the aging-related change rate of outcomes between HEMS pilots near the critical age of 60 and younger pilots, we allowed the slope of time since first available assessment to differ between “younger” and “older” pilots by including an interaction term between the binary age indicator and time. To maximize model parsimony, we did not include the age category main effect, assuming that any baseline age effects were accounted for by the continuous variable “age at first available assessment”. In order to assess the degree of unexplained heterogeneity between pilots relative to what can be explained by pilot age(ing), we compared change in Akaike’s Information Criterion (AIC) upon adding the fixed effects to the model after the random intercept to the corresponding change upon adding the random intercept after the fixed effects. We also calculated intraclass correlation coefficients (ICC), which quantify the degree to which variation is due to time-stable differences between persons rather than intraindividual change.(46) Finally, to give an overall impression of age-related change in predicted cardiovascular risk across the age range, we used a generalized additive model(47) to nonlinearly regress SCORE 6-month risk on age at examination separately in East and West European HEMS pilots, allowing for pilot-specific random intercepts.

2.5.3. Sensitivity analyses

To explore differences between East and West European HEMS pilots in the aging-related changes, we refitted all LMMs separately for these two pilot groups. We also repeated all LMM analyses including smoking status as a covariate.

2.5.4. Software

All analyses were carried out using R (version 3.3.2).(48) We used the packages nlme and mgcv, respectively, for linear mixed model and generalized additive model analyses.

3. RESULTS

3.1. Description of participants

Of 155 HEMS pilots targeted for participation, 75 (48.4%) agreed to participate. We obtained medical records for 66 of these (see Figure 1), 10 of whom had passed the age of 60 at their last available examination. The number of examinations available for each pilot ranged from 1 to 22 (mean: 14.8), with an average follow-up duration of 8.52 years (full details are provided in Supplementary Table SI). In general, follow-up was longer for East European compared to West European pilots, and the proportion of pilots aged 60 and over at the last available examination was much larger in the latter. Tables I and II give a descriptive impression of aging-related changes in auxiliary variables and risk marker levels, respectively.

3.2. Main results

The results of the LMM analyses are summarized in Table III. With regards to the main comparison of interest, that is, the difference in risk marker developments over follow-up time between “younger” and “older” pilots, the following pattern emerges: SBP changes little over time in younger pilots but appears to increase faster in the older pilots, although this finding is affected by large estimation uncertainty. QTc increases with aging in the younger pilots, and this increase appears to persist in the older pilots, although again with large estimation uncertainty. Developments in the remaining risk markers are in favor of the older pilots, as indicated by significant interaction effects: The aging-related increase in BMI and fasting glucose observable in younger pilots is notably reduced in older pilots, and their lipid profile improves over time, whereas it stays unchanged in the younger group. This improvement in cardiometabolic risk marker profile also impacts upon the SCORE risk development. Although it clearly increases with age – unsurprisingly, since SCORE risk values are deterministically calculated from age (and other risk markers) –, risk is estimated to increase by a factor of 1.09 per year in the older pilots, as opposed to 1.17 in the group of younger pilots. That is, the aging-related relative risk increase in older pilots is approximately half of that in younger pilots. ICCs reveal considerable time-stable differences between pilots in cardiometabolic risk markers (Table IV); 41% (fasting glucose) to 95% (BMI) of their variation is estimated to be due to unexplained between-pilot heterogeneity. In fact, accounting for unobserved time-stable pilot characteristics generally leads to much larger improvements in AIC model fit than age, aging, region, and medication taken together.

SCORE risks of fatal cardiovascular events within the next six months, as estimated by the generalized additive model, are shown in Figure 2 to give an impression of the overall relation between age and cardiovascular risk. Depending on age and region and under the assumption of no current blood pressure or lipid lowering drug use, risks are between 0 and 0.3%. Compared to SCORE risks of hypothetical individuals of the same age and region with an ideal modifiable risk factor profile (non-smoker, SBP 120 mmHg, TC 4 mmol/l) as a lower bound, estimated risks for HEMS pilots across the age range are about 50-100% larger. Since Poland and Czech Republic are classified as “high-risk” countries with a high cardiovascular mortality base rate, SCORE risk estimates are much higher for pilots from these countries, compared to German and Austrian pilots.

3.3. Results of sensitivity analyses

Including smoking as a covariate in the LMMs did not appreciably change the results presented in the previous section (Supplementary Table SII). Linear mixed models stratified by region suggest that SBP and QTc may particularly increase in older West European pilots (see Supplementary Table SIII). The favorable changes in BMI, HDLC, and fasting glucose are more pronounced in West, compared to East European older pilots, whereas the opposite is true for the favorable changes in TC. Given the fact that SCORE risk is partly determined by SBP and TC, these patterns may explain that the deceleration in relative SCORE risk increase in older pilots reported above is primarily present in the East European pilots in the stratified analysis. However, these results should be considered with great caution due to the, in part, extremely small sample sizes (see Supplementary Table SI).

4. LIMITATIONS

A major limitation of our study is its small sample size and the associated large estimation uncertainty, especially regarding the critical group of pilots aged 60 and over. This is partly due to the very existence of the Age 60 Rule itself, which makes it very hard to find active professional pilots aged 60 and over in the already small population of HEMS pilots. Data collection was further complicated by multiple stages of sample attrition: Several pilots did not respond to or declined our request for participation. The sensitive nature of the information involved, which is highly relevant to the pilots’ livelihood (continued certification of aeromedical fitness), likely is an important factor in this regard. Data then had to be obtained from many different centers and physicians, where again a sizable proportion (15 out of 38) did not respond. Moreover, the volume of received records strongly varied between sources; generally, aeromedical centers sent more extensive records. Data for East European pilots mainly came from such centers, whereas the data source for West European pilots were predominantly small private practices, explaining the follow-up time differences present in our data. In contrast, the number of active East European HEMS pilots aged 60 and over available to the study was much lower than initially expected, leading to disparate age distributions in East and West European pilots. Therefore, it was difficult to properly separate aging effects in East and West European pilots. In Europe, there is a persistent positive West-East gradient in cardiovascular disease burden,(49,50) but it is unclear to which degree this gradient also applies to professional pilots. Our stratified analyses suggested that the differences in risk marker developments between older and younger HEMS pilots were not homogeneous between the East and West European subgroups, although the small sample sizes involved prohibit more detailed or confident statements.

Finally, there was no possibility for us to confirm the validity of the findings documented in the examination records. The records came from different sources employing different procedures, materials, and documentation routines. Although we analyzed only well-established, easy-to-measure outcomes, it is possible that variation in measurement and documentation quality decreased the reliability of our results;(51) it is, however, not apparent that this would systematically bias the central comparison (aging-related change in markers in younger vs. older pilots).

5. DISCUSSION

In this longitudinal study, we investigated age-related changes in cardiometabolic risk markers among male European HEMS pilots to assess the validity of the “Age 60 Rule” in civil aviation regulations, which is – in the debate surrounding the rule – often justified on grounds of a possibly greater incapacitation risk in older pilots. We found that overall, compared to younger pilots, the development of the risk marker profile does not appear worse in pilots near age 60. In these older pilots, we observed a notable deceleration of aging-related BMI, fasting glucose, and cardiovascular risk score increases, and an improvement in lipid profile (total and HDL cholesterol) over time. Although affected by high estimation uncertainty, there was a tendency for systolic blood pressure to increase more rapidly over time in older pilots, and aging-related QTc interval increases already evident in the younger group appeared to persist. Large proportions of risk marker level variation were attributable to unexplained time-stable interindividual differences, rather than to observed variables. Fatal cardiovascular event risk within 6 months was estimated between 0 and 0.3%.

For several cardiometabolic risk factors (SBP, TC, fasting glucose, BMI), general population surveys have found an age-related increase in early and middle adulthood, which at some point in late adulthood appears to flatten or even reverse. In the case of TC and BMI, this may be as early as age 50-55, whereas it appears to occur later in the other risk factors, especially SBP.(52,53) The changes in developments seen in the older pilots in our sample may therefore simply reflect these trends. On the other hand, given that abnormal cardiovascular risk factor levels during aeromedical examinations usually lead to further diagnostic measures and in some cases to license revocations or restrictions (particularly, a requirement to fly multi-pilot only), a “healthy worker survivor” selection process(25,54) may also contribute to the beneficial changes observed in the older pilots.

Regardless of the reason for the observed age-related risk marker developments, they translate into a slowed increase in cardiovascular event risk. This is reflected in the slowed aging-related relative increase of the SCORE risk in the older HEMS pilots. Current European guidelines recommend the use of SCORE at ages 40-65,(37) which corresponds closely to the age range of interest here. Note that the deceleration is with regard to *relative* risk. In terms of *absolute* risk differences, the increase is larger in the older, compared to the younger pilots. However, the same is true when comparing, for example, a 40-year old to a 50-year old pilot; this argument therefore does not per se justify an age threshold of 60.

Another noteworthy aspect is the difference in estimated cardiovascular risk between West and East European pilots due to the classification of Poland and the Czech Republic as “high-risk” countries in current guidelines(37) because of their higher baseline levels of cardiovascular mortality. Whether this baseline difference also applies in the population of professional pilots is unknown. Furthermore, mortality differences between countries may not translate to corresponding differences in disease incidence, as case fatality rates may differ. A recently published cardiovascular risk score estimating fatal and nonfatal event risk calibrated to current baseline event rates in different countries suggests that 10-year event incidence in non-diabetic men aged 40-65 may actually be fairly similar in the four countries represented in our sample.(55)

Overall incidence is a more relevant outcome concerning flight safety than fatal event incidence, which underestimates the expected number of pilot incapacitations. On the other hand, available cardiovascular risk scores are calibrated to the general population, where incidence is higher than in professional pilots,(25,56) so that these biases tend to cancel each other out.(57) We chose SCORE as it allowed estimation of six-month risks instead of the usual ten-year risks, thus giving a more realistic approximation of the actual risk of cardiovascular incapacitation during the certificate validity period. The predicted risks, ranging between 0 and 0.3%, refer to in- and out of cockpit-time; the probability of an in-flight incapacitation event is correspondingly lower and additionally depends on the proportion of time spent flying.

For example, for a high-risk pilot (6-month predicted risk of 0.3%) flying on average 200 hours per half-year, the likelihood of an in-flight cardiovascular incapacitation during this half-year period would be estimated in-between 0.01 and 0.02%, given that a half-year approximately has 4,400 hours and assuming a constant hazard rate. More generally, in pilots with this risk level (which is at the very high end of our sample), the event rate would be 0.07 per 100,000 flight hours. Compared with the overall reported HEMS accident rates, for example 5.6 per 100,000 flight hours in the United States,(35) or 4.6 per 100,000 missions in Germany,(36) this seems to be a very minor factor; indeed, discussion of medical conditions as causes or contributory factors is virtually absent in reports on human factors issues in HEMS and commercial aviation accidents, and often do not feature at all in the included listings of causes (although it is hard to ascertain whether medical aspects may be subsumed under the listed categories).[2],(34,59)

Nevertheless, there is a strong sentiment among the public and regulators to take measures to control accident risk due to medical-cause pilot incapacitation – perhaps not least due to cognitive heuristics assigning a greater weight to abrupt, dramatic and seemingly uncontrollable events such as

---

[2] Also, where in-flight medical-cause incapacitations actually occur, most of the underlying conditions appear to be unrelated to aging (e.g., gastroenteritis, sinus conditions or headache).(58)

cardiovascular in-flight incapacitations than to, say, the much more common scenario of inadvertent flight into poor weather followed by spatial disorientation or loss of aircraft control.(60)

The issue of risk perception in aircraft accident causes that is mirrored in this sentiment had already been noted by Mason in 1974: “passengers have far less to fear from disease as a cause of aircraft accidents than might be supposed from the publicity which attend [sic] the subject”.(61) In the case of older professional pilots, such biases might be amplified by old-age stereotypes which still appear to be prevalent in the occupational context despite recent anti-discrimination efforts.(62) At the same time, beneficial effects of flight experience on accident risk(63,64) and the healthy worker survivor selection phenomenon appear to receive less attention. In our view, the most transparent, rational and fair way to resolve this issue would be to explicitly set levels of acceptable risk. The question of what level of risk is acceptable is, however, a normative problem. The weighting of the involved goods, such as flight safety and emergency response efficiency may be informed, but cannot be solved, by empirical findings alone.

The outcomes we studied are only surrogate endpoints for actual cardiovascular events. However, the rarity of these events in professional pilots makes their investigation challenging, even more so in the rather small subgroup of HEMS pilots. In 2004, 22 of 16,145 United Kingdom professional pilots (0.14%) suffered an incapacitating cardiovascular event or sudden death.(65) As of 2013, there were about 350 HEMS bases in Europe.(66) Assuming – very liberally – an average of six pilots per base, that is, about 2,100 active European HEMS pilots, 2.94 cardiovascular incapacitations would be expected to occur within one year. Moreover, no reports of actual HEMS accidents caused by pilot cardiovascular incapacitation appear to exist. Thus, investigating actual cardiovascular events in HEMS pilots would require a larger multinational effort (e.g., in the form of a case-control study). Within the scope of the present work, longitudinally assessing surrogate outcomes in the form of risk markers was therefore the only feasible option. Cardiovascular risk factors and scores were shown to be related to 5-year event incidence in professional pilots.(41) In the long run, the establishment of a cross-country harmonized database of aeromedical findings appears to be the only way to enable systematic investigations of incapacitating events in the small but healthy group of professional pilots.

6. CONCLUSION

To summarize, we found no indication of a worsening cardiometabolic risk marker profile in West and East European male HEMS pilots as they approached the age threshold for single-pilot commercial air transport operations of 60 years currently prescribed in the European Union (and other jurisdictions) compared to younger pilots. Our results therefore give no empirical support to the "Age 60 Rule", in line with the majority of earlier investigations of this long-discussed rule.(13) On the other hand, age clearly is a powerful predictor of cardiovascular events as it subsumes inexorable degenerative processes and cumulative exposures to risk factors like dyslipidemia.(67) The large time-stable interindividual variability in risk markers we found implies large differences in accumulated cardiovascular burden. To balance continued provision of comprehensive HEMS and optimal flight safety, we would therefore suggest as a short-term solution a one-off coronary artery calcium (CAC) scan at age 60 to guide the decision whether a pilot may continue flying single-pilot,(68) as this method captures cumulative effects of degenerative and pathological vascular processes and has a very high negative predictive value.(69–71) In the long term, to avoid unnecessary costs, radiation burden, and reduce false-positive findings, individualized risk scoring using an appropriately calibrated(72) prediction algorithm should be systematically applied well before age 60. Classical risk factor assessments should be considered longitudinally in terms of progression and accumulated burden, and, where appropriate, combined with modern diagnostic methods such as CAC. A longitudinal approach would be helpful in guiding preventive efforts and avoiding a single critical examination procedure which may result in a sudden forced grounding and which by itself acts as a job stressor that may threaten pilots' well-being, motivation, and flight safety.(73)

As others before us,(18,19) we thus recommend to eventually replace the general age threshold of 60 by a more individualized risk assessment with cardiovascular risk as only one (and likely a minor) aspect. Pilots already regularly need to undergo proficiency testing (and in some cases additionally line checks) to examine flight competency, and these could be modified and standardized for the specific case of the aging pilot to focus on cognitive and psychomotor capabilities known to decline on average with age, such as time sharing,(74) memory,(75) and performance speed.(76) However, the development of ecologically valid, standardized, and practicable performance testing procedures is likely to be a complex and resource-intensive endeavor. Specialized neuropsychological test batteries such as CogScreen-AE(77) may offer a feasible alternative, although in interpreting the results, it should be kept in mind that these might not fully capture the potentially beneficial effects older pilots' experience on actual flight performance.(78) Also, note that any type of performance testing for the purpose of selection will ultimately face the same issues of cutoff arbitrariness and sensitivity-specificity tradeoff as do cardiovascular risk scores, age, or any other pilot characteristics, for that matter. Independently of individual performance testing, a comprehensive risk management approach should also consider modifications of the operational environment to minimize accident risk (see, e.g., reference (34) for the case of HEMS).

Given the minor role of medical-cause sudden incapacitation in actual flight accidents, as well as the sensitivity-specificity-tradeoff noted above, one may even ask whether efforts invested into cardiovascular risk prediction in asymptomatic pilots were not better spent looking into the root causes of the more common accident scenarios, or focusing on actual flight performance and operational environment. Such a fundamental reorientation of risk management practice would, however, likely require a major shift in public thinking about risk. In the meantime, the development of a transparent, evidence-based cardiovascular risk assessment procedure, coupled with clearly defined acceptable risk levels, is a practicable solution that is less arbitrary than general age thresholds. The knowledge base and know-how acquired in the process could also prove beneficial to other occupational fields facing potential public safety risks associated with workforce aging.

REFERENCES

1. United Nations Department of Economic and Social Affairs Population Division. World Population Ageing 2015 (ST/ESA/SER.A/390). New York, NY: United Nations; 2015.
2. Dodman JM, Agius R, Turner S. Safety critical work, ageing and cognitive decline. Occup Med. 2012;62(7):480–2.
3. Chan G, Tan V, Koh D. Ageing and fitness to work. Occup Med. 2000;50(7):483–91.
4. Sluiter JK. High-demand jobs: age-related diversity in work ability? Appl Ergon. 2006;37(4):429–40.

5. Hack MR. Taking age equality seriously: the example of mandatory retirement. A comparative legal analysis between Norway and Germany in light of Council Directive 2000/78/EC. 1st ed. Baden-Baden: Nomos; 2016.
6. Woerth DE. Statement of Captain Duane E. Woerth, President, Air Line Pilots Association, before the Committee on Commerce, Science and Transportation, United States Senate, on the “Age 60 Rule”. United States Senate. 2001 Mar 13. Available from: https://www.gpo.gov/fdsys/pkg/CHRG-109shrg63516/html/CHRG-109shrg63516.htm
7. Wilkening R. The Age 60 Rule: age discrimination in commercial aviation. Aviat Space Environ Med. 2002;73(3):194–202.
8. Lawrence PA. Retiring retirement. Nature. 2008;453(7195):588–90.
9. Powell JL, Taylor P. Rethinking risk and ageing. extending working lives. Soc Policy Soc. 2016;15(4):637–45.
10. Organisation for Economic Cooperation and Development. Live longer, work longer. Paris: OECD; 2006.
11. Dellinger EP, Pellegrini CA, Gallagher TH. The aging physician and the medical profession: a review. JAMA Surg. 2017;152(10):967–71.
12. Sachverständigenkommission zur Erstellung des Sechsten Altenberichts der Bundesregierung. Sechster Bericht zur Lage der älteren Generation in der Bundesrepublik Deutschland – Altersbilder in der Gesellschaft [Sixth report on the situation of the older generation in the Federal Republic of Germany – images of old age in society]. Berlin: Bundesministerium für Familie, Senioren, Frauen und Jugend; 2010.
13. Aerospace Medical Association. The Age 60 Rule. Aviat Space Environ Med. 2004;75(8):708–15.
14. Cornell A, Baker SP, Li G. Age-60 Rule: the end is in sight. Aviat Space Environ Med. 2007;78(6):624–6.
15. Skirbekk V, Loichinger E, Barakat BF. The aging of the workforce in European countries: demographic trends, retirement projections, and retirement policies. In: Hedge JW, Borman

WC, editors. The Oxford handbook of work and aging. Oxford: Oxford University Press; 2012. p. 60–79.

16. Ardila A. Normal aging increases cognitive heterogeneity: analysis of dispersion in WAIS-III scores across age. Arch Clin Neuropsychol. 2007;22(8):1003–11.
17. Lowsky DJ, Olshansky SJ, Bhattacharya J, Goldman DP. Heterogeneity in healthy aging. J Gerontol A Biol Sci Med Sci. 2014;69(6):640–9.
18. Stuck AE, van Gorp WG, Josephson KR, Morgenstern H, Beck JC. Multidimensional risk assessment versus age as criterion for retirement of airline pilots. J Am Geriatr Soc. 1992;40(5):526–32.
19. Birren JE, Fisher LM. Rules and reason in the forced retirement of commercial airline pilots at age 60. Ergonomics. 1995;38(3):518–25.
20. Evans A. Aeromedical risk: a numerical approach. In: Rainford D, Gradwell DP, editors. Ernsting's aviation and space medicine. Boca Raton, FL: CRC Press; 2016. p. 373–84.
21. Houston S, Mitchell S, Evans S. Application of a cardiovascular disease risk prediction model among commercial pilots. Aviat Space Environ Med. 2010;81(8):768–73.
22. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Despres JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Liu S, Mackey RH, Matchar DB, McGuire DK, Mohler ER, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Nichol G, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner MB. Heart disease and stroke statistics - 2015 update: a report from the American Heart Association. Circulation. 2015;131(4):e29–322.
23. Hardy R, Lawlor DA, Kuh D. A life course approach to cardiovascular aging. Future Cardiol. 2015;11(1):101–13.
24. Institute for Health Metrics and Evaluation (IHME). GBD Compare data visualization. Seattle, WA: IHME; 2016. Available from: http://ihmeuw.org/3zrg

25. Linnersjö A, Brodin LA, Andersson C, Alfredsson L, Hammar N. Low mortality and myocardial infarction incidence among flying personnel during working career and beyond. Scand J Work Environ Health. 2011;37(3):219–26.

26. Hammer GP, Auvinen A, de Stavola BL, Grajewski B, Gundestrup M, Haldorsen T, Hammar N, Lagorio S, Linnersjö A, Pinkerton L, Pukkala E, Rafnsson V, dos-Santos-Silva I, Storm HH, Strand TE, Tzonou A, Zeeb H, Blettner M. Mortality from cancer and other causes in commercial airline crews: a joint analysis of cohorts from 10 countries. Occup Environ Med. 2014;71(5):313–22.

27. Huster KM, Müller A, Prohn MJ, Nowak D, Herbig B. Medical risks in older pilots: a systematic review on incapacitation and age. Int Arch Occup Environ Health. 2014;87(6):567–78.

28. Poguntke P. Interview with F. Bruder, new CEO of ADAC Air Rescue: “Highest medical quality for people in distress”. AirRescue Magazine. 2012;2(4):14–5.

29. Hickman BJ, Mehrer R. Stress and the effects of air transport on flight crews. Air Med J. 2001;20(6):6–9.

30. Carchietti E, Valent F, Cecchi A, Rammer R. Influence of stressors on HEMS crewmembers in flight. Air Med J. 2011;30(5):270–5.

31. Radstaak M, Geurts SA, Beckers DG, Brosschot JF, Kompier MA. Work stressors, perseverative cognition and objective sleep quality: a longitudinal study among Dutch Helicopter Emergency Medical Service (HEMS) Pilots. J Occup Health. 2014;56(6):469–77.

32. Samel A, Vejvoda M, Maass H. Sleep deficit and stress hormones in helicopter pilots on 7-day duty for emergency medical services. Aviat Space Environ Med. 2004;75(11):935–40.

33. Rodenberg H, Blumen IJ, Thomas SH. Air medical transport. In: Marx JA, Hockberger RS, Walls RM, editors. Rosen’s emergency medicine, Concepts and clinical practice. Philadelphia, PA: Elsevier Saunders; 2014. p. 2442–48.e2.

34. Blumen IJ, UCAN Safety Committee. A safety review and risk assessment in air medical transport: supplement to the Air Medical Physician Handbook. Salt Lake City, UT: Air Medical Physician Association; 2002.

35. Blumen I. An analysis of HEMS accidents and accident rates. Presentation at NTSB public hearing on the safety of helicopter emergency medical services; 2009 Feb 3. Available from: https://www.ntsb.gov/news/events/Pages/2009-HEMS_agenda.aspx

36. Hinkelbein J, Schwalbe M, Neuhaus C, Wetsch WA, Genzwürker HV. Incidents, accidents and fatalities in 40 years of German helicopter emergency medical system operations. Eur J Anaesthesiol. 2011;28(11):766–73.

37. Piepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, Catapano AL, Cooney MT, Corra U, Cosyns B, Deaton C, Graham I, Hall MS, Hobbs FD, Lochen ML, Lollgen H, Marques-Vidal P, Perk J, Prescott E, Redon J, Richter DJ, Sattar N, Smulders Y, Tiberi M, van der Worp, H Bart, van Dis I, Verschuren WM. 2016 European Guidelines on cardiovascular disease prevention in clinical practice. Eur Heart J. 2016;37(29):2315–81.

38. Tabák AG, Herder C, Rathmann W, Brunner EJ, Kivimaki M. Prediabetes: a high-risk state for diabetes development. Lancet. 2012;379(9833):2279–90.

39. Eckel RH, York DA, Rossner S, van Hubbard, Caterson I, St Jeor ST, Hayman LL, Mullis RM, Blair SN. Prevention Conference VII: Obesity, a worldwide epidemic related to heart disease and stroke: executive summary. Circulation. 2004;110(18):2968–75.

40. Straus SM, Kors JA, de Bruin ML, van der Hooft CS, Hofman A, Heeringa J, Deckers JW, Kingma JH, Sturkenboom MC, Stricker BH, Witteman JC. Prolonged QTc interval and risk of sudden cardiac death in a population of older adults. J Am Coll Cardiol. 2006;47(2):362–7.

41. Wirawan IM, Larsen PD, Aldington S, Griffiths RF, Ellis CJ. Cardiovascular risk score and cardiovascular events among airline pilots: a case-control study. Aviat Space Environ Med. 2012;83(5):465–71.

42. Evans A, Evans S, Harper G. International regulation of medical standards. In: Rainford D, Gradwell DP, editors. Ernsting's aviation and space medicine. Boca Raton, FL: CRC Press; 2016. p. 357–72.

43. Müller A, Prohn MJ, Huster KM, Nowak D, Angerer P, Herbig B. Pilots' age and incidents in helicopter emergency medical services: a 5-year observational study. Aviat Space Environ Med. 2014;85(5):522–8.

44. Conroy RM, Pyörälä K, Fitzgerald AP, Sans S, Menotti A, de Backer G, de Bacquer D, Ducimetière P, Jousilahti P, Keil U, Njølstad I, Oganov RG, Thomsen T, Tunstall-Pedoe H, Tverdal A, Wedel H, Whincup P, Wilhelmsen L, Graham IM. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. Eur Heart J. 2003;24(11):987–1003.

45. Pinheiro JC, Bates DM. Mixed-effects models in S and S-PLUS. New York, NY: Springer; 2000.

46. Bliese PD, Wesensten NJ, Balkin TJ. Age and individual variability in performance during sleep restriction. J Sleep Res. 2006;15(4):376–85.

47. Wood SN. Generalized additive models: An introduction with R. Boca Raton, FL: CRC Press; 2006.

48. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2016.

49. Moran AE, Forouzanfar MH, Roth GA, Mensah GA, Ezzati M, Flaxman A, Murray CJ, Naghavi M. The global burden of ischemic heart disease in 1990 and 2010: the Global Burden of Disease 2010 study. Circulation. 2014;129(14):1493–501.

50. Müller-Nordhorn J, Binting S, Roll S, Willich SN. An update on regional variation in cardiovascular mortality within Europe. Eur Heart J. 2008;29(10):1316–26.

51. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc. 2013;20(1):144–51.

52. Wills AK, Lawlor DA, Matthews FE, Sayer AA, Bakra E, Ben-Shlomo Y, Benzeval M, Brunner E, Cooper R, Kivimaki M, Kuh D, Muniz-Terrera G, Hardy R. Life course trajectories of systolic blood pressure using longitudinal data from eight UK cohorts. PLoS Med. 2011;8(6):e1000440.

53. Singh GM, Danaei G, Pelizzari PM, Lin JK, Cowan MJ, Stevens GA, Farzadfar F, Khang YH, Lu Y, Riley LM, Lim SS, Ezzati M. The age associations of blood pressure, cholesterol, and glucose: analysis of health examination surveys from international populations. Circulation. 2012;125(18):2204–11.

54. Arrighi HM, Hertz-Picciotto I. The evolving concept of the healthy worker survivor effect. Epidemiology. 1994;5(2):189–96.

55. Ueda P, Woodward M, Lu Y, Hajifathalian K, Al-Wotayan R, Aguilar-Salinas CA, Ahmadvand A, Azizi F, Bentham J, Cifkova R, Di Cesare M, Eriksen L, Farzadfar F, Ferguson TS, Ikeda N, Khalili D, Khang YH, Lanska V, Leon-Munoz L, Magliano DJ, Margozzini P, Msyamboza KP, Mutungi G, Oh K, Oum S, Rodriguez-Artalejo F, Rojas-Martinez R, Valdivia G, Wilks R, Shaw JE, Stevens GA, Tolstrup JS, Zhou B, Salomon JA, Ezzati M, Danaei G. Laboratory-based and office-based risk scores and charts to predict 10-year risk of cardiovascular disease in 182 countries: a pooled analysis of prospective cohorts and health surveys. Lancet Diabetes Endocrinol. 2017;5(3):196–213.

56. de Stavola BL, Pizzi C, Clemens F, Evans SA, Evans AD, dos Santos Silva I. Cause-specific mortality in professional flight crew and air traffic control officers: findings from two UK population-based cohorts of over 20,000 subjects. Int Arch Occup Environ Health. 2012;85(3):283–93.

57. Mitchell SJ, Evans AD. Flight safety and medical incapacitation risk of airline pilots. Aviat Space Environ Med. 2004;75(3):260–8.

58. James M, Green R. Airline pilot incapacitation survey. Aviat Space Environ Med. 1991;62(11):1068–72.

59. European Helicopter Safety Team. EHEST Analysis of 2006 – 2010 European Helicopter Accidents. Cologne: European Aviation Safety Agency; 2015.

60. Butler B. Helicopter emergency medical services and weather-related accidents. Air Med J. 2014;33(3):84–5.

61. Mason JK. Disease of aircrew as a cause of aircraft accidents. Community Health. 1974;6(2):62–8.

62. Harris K, Krygsman S, Waschenko J, Laliberte Rudman D. Ageism and the older worker: a scoping review. Gerontologist. 2017. doi:10.1093/geront/gnw194

63. Li G, Baker SP, Grabowski JG, Qiang Y, McCarthy ML, Rebok GW. Age, flight experience, and risk of crash involvement in a cohort of professional pilots. Am J Epidemiol. 2003;157(10):874–80.

64. Aherne BB, Zhang C, Newman DG. Pilot domain task experience in night fatal helicopter emergency medical service accidents. Aerosp Med Hum Perform. 2016;87(6):550–6.

65. Evans S, Radcliffe SA. The annual incapacitation rate of commercial pilots. Aviat Space Environ Med. 2012;83(1):42–9.

66. Kessler C. Helicopter emergency medical service: motivation for focused research. CEAS Aeronautical Journal. 2015;6(3):337–94.

67. Sniderman AD, Furberg CD. Age as a modifiable risk factor for cardiovascular disease. Lancet. 2008;371(9623):1547–9.

68. Wirawan IM, Wu R, Abernethy M, Aldington S, Larsen PD. Calcium scores in the risk assessment of an asymptomatic population: implications for airline pilots. Aviat Space Environ Med. 2014;85(8):812–7.

69. Sarwar A, Shaw LJ, Shapiro MD, Blankstein R, Hoffmann U, Cury RC, Abbara S, Brady TJ, Budoff MJ, Blumenthal RS, Nasir K. Diagnostic and prognostic value of absence of coronary artery calcification. JACC Cardiovasc Imaging. 2009;2(6):675–88.

70. Valenti V, ó Hartaigh B, Heo R, Cho I, Schulman-Marcus J, Gransar H, Truong QA, Shaw LJ, Knapper J, Kelkar AA, Sandesara P, Lin FY, Sciarretta S, Chang HJ, Callister TQ, Min JK. A 15-year warranty period for asymptomatic individuals without coronary artery calcium: a prospective follow-up of 9,715 individuals. JACC Cardiovasc Imaging. 2015;8(8):900–9.

71. Blaha MJ, Cainzos-Achirica M, Greenland P, McEvoy JW, Blankstein R, Budoff MJ, Dardari Z, Sibley CT, Burke GL, Kronmal RA, Szklo M, Blumenthal RS, Nasir K. Role of coronary artery calcium score of zero and other negative risk markers for cardiovascular disease: the Multi-Ethnic Study of Atherosclerosis (MESA). Circulation. 2016;133(9):849–58.

72. Blaha MJ. The critical importance of risk score calibration: time for transformative approach to risk score validation? J Am Coll Cardiol. 2016;67(18):2131–4.

73. Sverke M, Hellgren J, Näswall K. No security: a meta-analysis and review of job insecurity and its consequences. J Occup Health Psychol. 2002;7(3):242–64.

74. Verhaeghen P, Steitz DW, Sliwinski MJ, Cerella J. Aging and dual-task performance: a meta-analysis. Psychol Aging. 2003;18(3):443–60.

75. Nilsson LG. Memory function in normal aging. Acta Neurol Scand Suppl. 2003;179:7–13.

76. Desrocher M, Smith ML. Intrapersonal and extrapersonal space: performance of older adults on ecologically valid orientation tasks. Exp Aging Res. 2005;31(2):205–16.

77. Taylor JL, O’Hara R, Mumenthaler MS, Yesavage JA. Relationship of CogScreen-AE to flight simulator performance and pilot age. Aviat Space Environ Med. 2000;71(4):373–80.

78. Hardy DJ, Parasuraman R. Cognition and flight performance in older pilots. J Exp Psychol Appl. 1997;3(4):313–48.

TABLES

Table I. Descriptives for binary variables and mean age at first and last available assessment for pilots with ≥2 assessments of the respective variable, by region of Europe and pilot age

		East[a]									**West**[a]									**Total**								
		Younger[b]			**Older**[b]			**Total**			**Younger**[b]			**Older**[b]			**Total**			**Younger**[b]			**Older**[b]			**Total**		
		n	%	$\overline{\text{Age}}$	n	%	$\overline{\text{Age}}$	n	%	$\overline{\text{Age}}$	n	%	$\overline{\text{Age}}$	n	%	$\overline{\text{Age}}$	n	%	$\overline{\text{Age}}$	n	%	$\overline{\text{Age}}$	n	%	$\overline{\text{Age}}$	n	%	$\overline{\text{Age}}$
Smoking	First	11	0.28	39.6	1	0.50	53.0	12	0.29	40.2	1	0.07	46.2	1	0.14	55.8	2	0.09	49.3	12	0.22	41.4	2	0.22	55.1	14	0.22	43.3
	Last	12	0.30	48.5	0	0.00	62.1	12	0.29	49.2	1	0.07	54.1	0	0.00	61.1	1	0.05	56.3	13	0.24	50.0	0	0.00	61.3	13	0.20	51.6
BP lowering drugs	First	1	0.03	41.6	0	0.00	53.0	1	0.02	42.2	1	0.07	46.2	0	0.00	55.8	1	0.05	49.3	2	0.04	42.9	0	0.00	55.1	2	0.03	44.6
	Last	4	0.10	49.1	1	0.50	62.3	5	0.12	49.7	3	0.20	54.1	1	0.14	61.1	4	0.18	56.3	7	0.13	50.5	2	0.22	61.3	9	0.14	52.0
Lipid lowering drugs	First	1	0.03	41.6	0	0.00	53.0	1	0.02	42.2	0	0.00	46.2	0	0.00	55.8	0	0.00	49.3	1	0.02	42.9	0	0.00	55.1	1	0.02	44.6
	Last	4	0.10	49.1	0	0.00	62.3	4	0.10	49.7	1	0.07	54.1	0	0.00	61.1	1	0.05	56.3	5	0.09	50.5	0	0.00	61.3	5	0.08	52.0

BP: Blood pressure. $\overline{\text{Age}}$: Mean age. [a]"East"/"West": Czech Republic, Poland/Germany, Austria. [b]"Younger"/"Older": Aged <60/≥60 yrs. at last available examination.

Table II. Descriptives for quantitative variables and mean age at first and last available assessment for pilots with ≥2 assessments of the respective variable, by region of Europe and pilot age

		East[a]									**West**[a]									**Total**								
		Younger[b]			**Older**[b]			**Total**			**Younger**[b]			**Older**[b]			**Total**			**Younger**[b]			**Older**[b]			**Total**		
		M	SD	$\overline{\text{Age}}$	M	SD	$\overline{\text{Age}}$	M	SD	$\overline{\text{Age}}$	M	SD	$\overline{\text{Age}}$	M	SD	$\overline{\text{Age}}$	M	SD	$\overline{\text{Age}}$	M	SD	$\overline{\text{Age}}$	M	SD	$\overline{\text{Age}}$	M	SD	$\overline{\text{Age}}$
SBP [mmHg]	First	128	11	39.7	138	4	53.0	128	11	40.3	130	11	46.8	123	11	55.8	128	11	49.6	128	11	41.7	126	12	55.1	128	11	43.6
	Last	131	13	48.5	134	6	62.3	131	12	49.2	129	10	53.7	126	15	61.1	128	11	55.9	131	12	50.0	128	14	61.3	130	12	51.6
QTc [ms]	First	393	15	38.6	375	---	52.7	392	15	39.1	406	28	46.9	406	5	56.5	406	24	49.3	397	21	41.3	400	15	55.7	397	20	43.0
	Last	403	23	46.7	422	---	63.2	404	23	47.4	406	18	55.1	415	17	60.8	408	18	56.5	404	21	49.4	416	15	61.3	406	21	50.8
BMI [kg/m²]	First	26.4	3.3	39.6	29.5	4	53.4	26.6	3.4	40.3	25.8	3	46.6	27.8	2.5	55.8	26.4	2.9	49.4	26.3	3.2	41.7	28.2	2.7	55.2	26.5	3.2	43.6
	Last	27.4	3.4	48.7	30	2.8	62.3	27.6	3.4	49.4	26.3	3.5	53.7	27.6	2.6	61.1	26.7	3.3	55.9	27.1	3.4	50.2	28.1	2.7	61.3	27.3	3.4	51.7
TC [mmol/l]	First	5.3	1.1	40.3	5.5	0.8	53.5	5.3	1.1	41.0	5.7	0.7	46.8	4.7	0.5	54.2	5.4	0.7	48.7	5.4	1.0	41.8	5.0	0.7	54.0	5.4	1.0	43.1
	Last	5.2	0.9	48.5	5.3	0.5	62.3	5.2	0.8	49.1	5.5	0.8	52.9	4.8	0.5	60.2	5.4	0.8	54.7	5.3	0.8	49.5	4.9	0.5	60.9	5.2	0.8	50.7
HDLC [mmol/l]	First	1.4	0.2	40.8	1.8	---	53.2	1.4	0.3	41.2	1.4	0.4	46.3	1.5	0.3	55.7	1.4	0.4	48.3	1.4	0.3	42.1	1.6	0.3	55.0	1.4	0.3	43.2
	Last	1.4	0.4	47.4	2.0	---	61.5	1.4	0.4	47.8	1.3	0.3	52.3	1.8	0.2	61.1	1.4	0.4	54.2	1.4	0.3	48.6	1.9	0.2	61.2	1.4	0.4	49.6
Glu [mmol/l]	First	5.3	0.5	40.5	5.3	0.1	53.5	5.3	0.5	41.1	4.8	1.1	47.6	5.4	0.7	54.0	4.9	1.0	49.3	5.2	0.7	42.0	5.3	0.5	53.8	5.2	0.7	43.3
	Last	5.3	0.4	48.3	5.7	1.0	62.3	5.3	0.5	48.9	4.7	1.0	52.2	4.8	0.4	60.2	4.7	0.9	54.3	5.2	0.6	49.1	5.1	0.7	60.9	5.2	0.6	50.3
SCORE	First	0.03	0.04	40.4	0.1	0.02	53.5	0.03	0.04	41.0	0.02	0.02	47.0	0.04	0.03	55.5	0.03	0.02	49.3	0.03	0.04	41.8	0.06	0.03	54.8	0.03	0.04	43.2
(6 months) [%][c]	Last	0.07	0.07	48.5	0.2	0.02	62.3	0.08	0.07	49.1	0.05	0.02	53.5	0.07	0.04	58.9	0.05	0.03	54.9	0.07	0.06	49.6	0.11	0.07	60.1	0.07	0.06	50.7

M: Mean. SD: Standard deviation. $\overline{\text{Age}}$: Mean age. SBP: Systolic blood pressure. QTc: Corrected QT interval duration. BMI: Body mass index. TC: Total cholesterol. HDLC: High-density lipoprotein cholesterol. Glu: Fasting glucose. SCORE: Systematic Coronary Risk Evaluation. ---: One individual only, standard deviation not estimable. [a]"East"/"West": Czech Republic, Poland/Germany, Austria. [b]"Younger"/"Older": Aged <60/≥60 yrs. at last available examination. [c]Probability to suffer a fatal cardiovascular event within 6 months as estimated by the SCORE algorithm.

Table III. Linear mixed model coefficients for cardiometabolic risk factors and fatal cardiovascular disease risk prediction in HEMS pilots.

		SBP [mmHg]		**QTc [ms]**		**BMI [kg/m^2]**		**TC [mmol/l]**		**HDLC [mmol/l]**		**Glu [mmol/l]**		**SCORE (6 m.) [%][a,b]**	
		Coef.	**95% CI**	**Coef.**	**95% CI**	**Coef.**	**95% CI**	**Coef.**	**95% CI**	**Coef.**	**95% CI**	**Coef.**	**95% CI**	**Coef.**	**95% CI**
Intercept[c]		126.6	122.7, 130.6	402.2	393.1, 411.3	25.3	23.7, 26.9	5.29	4.87, 5.71	1.25	1.08, 1.41	4.7	4.45, 4.95	0.004	0.003, 0.005
East European[d]		3.48	-1.05, 8.00	-10.37	-21.58, 0.83	1.28	-0.63, 3.19	0.01	-0.46, 0.47	0.19[h]	0.01, 0.38	0.45[h]	0.18, 0.72	2.78[h]	2.14, 3.62
Relevant medication[e]		-0.70	-3.92, 2.53	---	---	---	---	-1.35[h]	-1.60, -1.10	-0.06	-0.16, 0.05	---	---	0.73[h]	0.66, 0.81
Age at first available assessment [yrs.][f]		0.01	-0.25, 0.27	0.36	-0.26, 0.98	0.09	-0.02, 0.19	0.02	-0.01, 0.04	0.02[h]	0.00, 0.03	0.02[h]	0.00, 0.03	1.19[h]	1.18, 1.21
Time since first available assessment [yrs.]	Younger[g]	0.05	-0.14, 0.24	1.2[h]	0.75, 1.69	0.12[h]	0.10, 0.14	-0.01	-0.02, 0.01	-0.01	-0.01, 0.00	0.02[h]	0.00, 0.03	1.17[h]	1.16, 1.18
	Older[g]	0.21	-0.38, 0.79	1.34	-0.48, 3.15	0.03	-0.04, 0.08	-0.07[h]	-0.13, -0.02	0.04[h]	0.01, 0.06	-0.03	-0.07, 0.01	1.09[h]	1.05, 1.12
	Difference[g]	0.16	-0.45, 0.76	0.12	-1.75, 1.98	-0.10[h]	-0.16, -0.04	-0.07[h]	-0.12, -0.01	0.04[h]	0.01, 0.07	-0.05[h]	-0.09, -0.01	0.93[h]	0.90, 0.95

SBP: Systolic blood pressure. QTc: Corrected QT interval duration. BMI: Body mass index. TC: Total cholesterol. HDLC: High-density lipoprotein cholesterol. Glu: Fasting glucose. SCORE: Systematic Coronary Risk Evaluation. CI: Confidence interval. [a]Probability to suffer a fatal cardiovascular event within 6 months as estimated by the SCORE algorithm. [b]Coefficients are multiplicative for this outcome (vs. additive for all others). [c]40-yr. old West European not currently using relevant medication, at first available assessment. [d]Pilot from Czech Republic or Poland (vs. Germany or Austria). [e]SBP: Blood-pressure lowering drugs, TC, HDLC: Lipid-lowering drugs, SCORE: Either type of drug. [f]Centered at 40 years. [g]“Younger”/“Older”: Time effect for pilots aged <60/≥60 yrs. at last available examination, “Difference”: Difference between these effects. [h]95% CI for covariate effect does not include null value (1 for SCORE, 0 for all other outcomes).

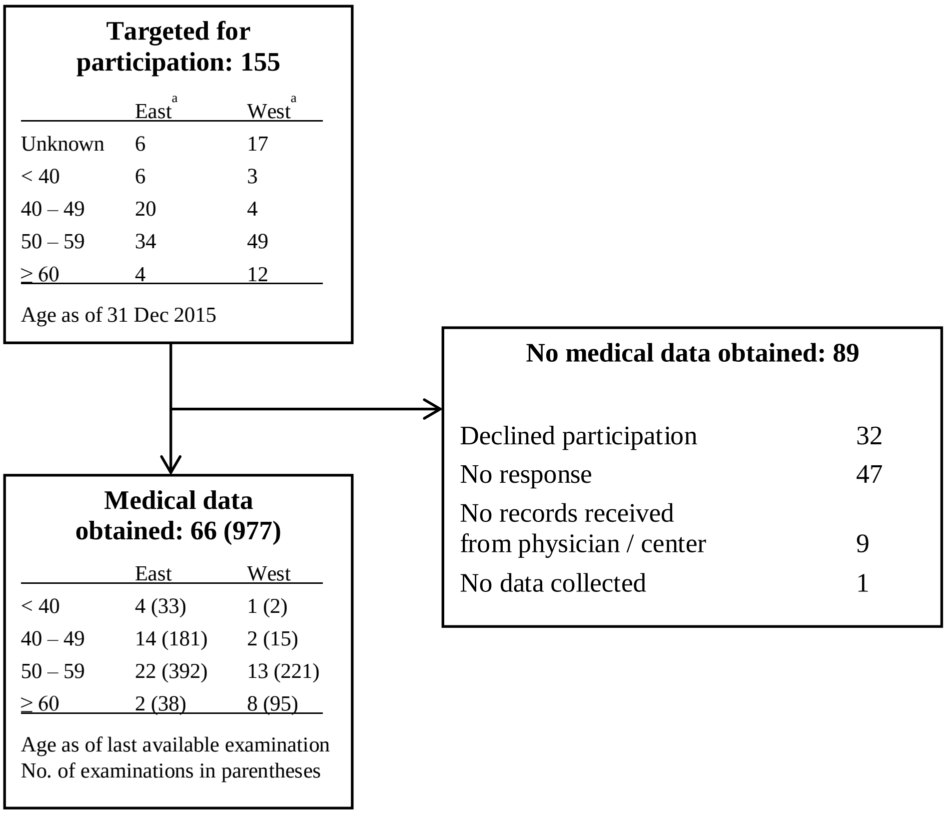
Table IV. Linear mixed model measures of unexplained heterogeneity for cardiometabolic risk factors and fatal cardiovascular disease risk prediction in HEMS pilots.

		SBP [mmHg]	**QTc [ms]**	**BMI [kg/m²]**	**TC [mmol/l]**	**HDLC [mmol/l]**	**Glu [mmol/l]**	**SCORE (6 m.) [%][a]**
ΔAIC[b]	All fixed effects added to random intercept	+6.0	-26.6	-158.0	-100.4	-12.3	-12.0	-899.7
	Random intercept added to all fixed effects	-301.7	-126.1	-2,184.8	-385.4	-357.6	-136.9	-372.9
ICC [%]		44.8	45.5	94.8	61.9	68.8	40.5	68.3

SBP: Systolic blood pressure. QTc: Corrected QT interval duration. BMI: Body mass index. TC: Total cholesterol. HDLC: High-density lipoprotein cholesterol. Glu: Fasting glucose. SCORE: Systematic Coronary Risk Evaluation. AIC: Akaike's Information Criterion. ICC: Intraclass coefficient.

[a] Probability to suffer a fatal cardiovascular event within 6 months as estimated by the SCORE algorithm. [b] Based on maximum likelihood estimation.

FIGURES



Targeted for participation: 155

	East[a]	West[a]
Unknown	6	17
< 40	6	3
40 – 49	20	4
50 – 59	34	49
≥ 60	4	12

Age as of 31 Dec 2015

No medical data obtained: 89

Declined participation	32
No response	47
No records received from physician / center	9
No data collected	1

Medical data obtained: 66 (977)

	East	West
< 40	4 (33)	1 (2)
40 – 49	14 (181)	2 (15)
50 – 59	22 (392)	13 (221)
≥ 60	2 (38)	8 (95)

Age as of last available examination
No. of examinations in parentheses

Fig. 1. Study participation by age and region of origin. [a]"East"/"West": Czech Republic, Poland/Germany, Austria

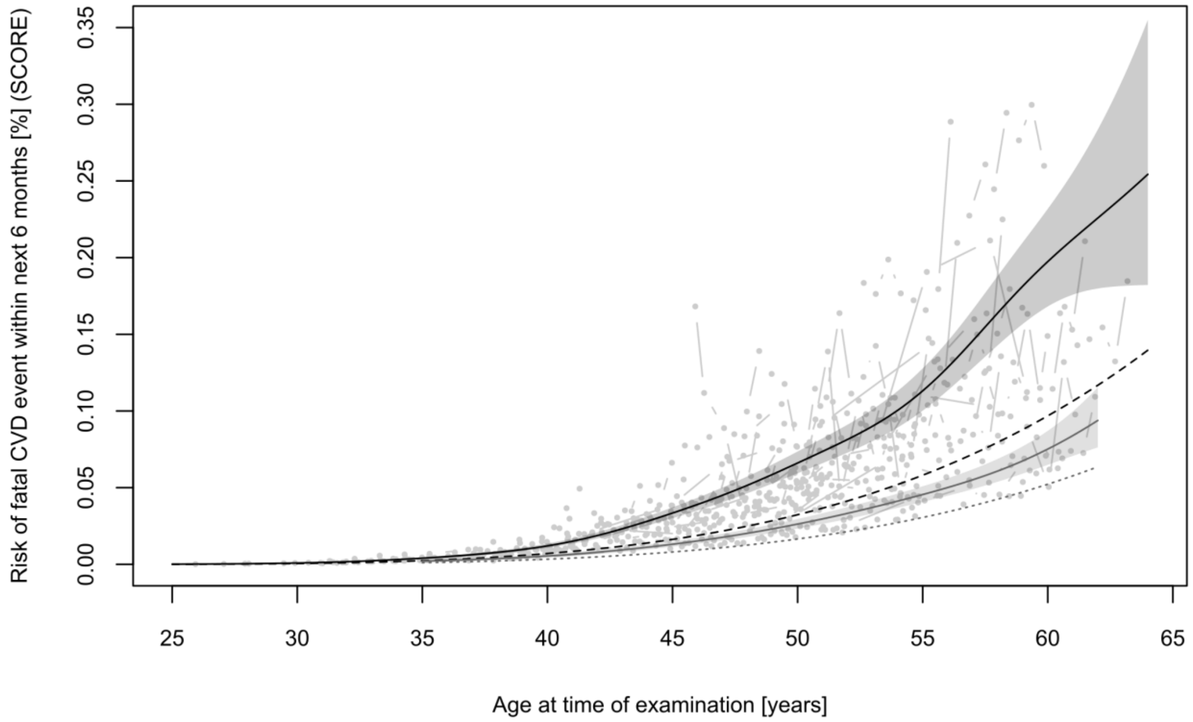


Fig. 2. SCORE risk of fatal cardiovascular event within six months, by age. Solid lines: GAM estimates (with 95% pointwise confidence intervals) for East European (Polish, Czech; dark shade) and West European (Austrian, German; light shade) pilots not currently using blood pressure or lipid lowering drugs. Dashed/dotted line: SCORE risk of man from same country with ideal modifiable cardiovascular risk profile (SBP 120 mmHg, TC 4 mmol/l, nonsmoker). Actual data points in background as dots with interpolating lines between measurements in same individual.

SUPPLEMENTARY MATERIAL

Supplementary Table SI. Number of subjects, number of observations, and average follow-up duration by region of Europe, and pilot age

			East[a]			**West[a]**			**Total**		
			Younger[b]	**Older[b]**	**Total**	**Younger**	**Older**	**Total**	**Younger**	**Older**	**Total**
Smoking	≥1 available assessments[c]	Avg. FU	8.94	9.13	8.95	7.86	4.63	6.74	8.65	5.53	8.17
		# Obs.	573	33	606	221	74	295	794	107	901
		# Subj.	40	2	42	15	8	23	55	10	65
	≥2 available assessments[c]	Avg. FU	8.94	9.13	8.95	7.86	5.29	7.04	8.65	6.15	8.29
		# Obs.	573	33	606	221	73	294	794	106	900
		# Subj.	40	2	42	15	7	22	55	9	64
Blood pressure lowering drugs	≥1 available assessments	Avg. FU	7.26	9.38	7.36	7.86	4.63	6.74	7.42	5.58	7.14
		# Obs.	483	35	518	221	75	296	704	110	814
		# Subj.	40	2	42	15	8	23	55	10	65
	≥2 available assessments	Avg. FU	7.45	9.38	7.54	7.86	5.29	7.04	7.56	6.20	7.37
		# Obs.	482	35	517	221	74	295	703	109	812
		# Subj.	39	2	41	15	7	22	54	9	63
Lipid lowering drugs	≥1 available assessments	Avg. FU	7.26	9.38	7.36	7.86	4.63	6.74	7.42	5.58	7.14
		# Obs.	483	35	518	221	75	296	704	110	814
		# Subj.	40	2	42	15	8	23	55	10	65
	≥2 available assessments	Avg. FU	7.45	9.38	7.54	7.86	5.29	7.04	7.56	6.20	7.37
		# Obs.	482	35	517	221	74	295	703	109	812
		# Subj.	39	2	41	15	7	22	54	9	63
SBP	≥1 available assessments	Avg. FU	8.84	9.38	8.87	6.82	4.63	6.09	8.26	5.58	7.86
		# Obs.	574	35	609	196	67	263	770	102	872
		# Subj.	40	2	42	16	8	24	56	10	66
	≥2 available assessments	Avg. FU	8.84	9.38	8.87	6.82	5.29	6.36	8.26	6.20	7.98
		# Obs.	574	35	609	196	66	262	770	101	871
		# Subj.	40	2	42	16	7	23	56	9	65

SBP: Systolic blood pressure. QTc: Corrected QT interval duration. BMI: Body mass index. TC: Total cholesterol. HDLC: High-density lipoprotein cholesterol. Glu: Fasting glucose. SCORE: Systematic Coronary Risk Evaluation. FU: Follow-up in years. [a]"East"/"West": Czech Republic, Poland/Germany, Austria. [b]"Younger"/"Older": Aged <60/≥60 yrs. at last available examination. [c]All pilots with at least one/two available assessment(s) of respective outcome.

Supplementary Table SI. Number of subjects, number of observations, and average follow-up duration by region of Europe, and pilot age

			East[a]			**West[a]**			**Total**		
			Younger[b]	**Older[b]**	**Total**	**Younger**	**Older**	**Total**	**Younger**	**Older**	**Total**
QTc	≥1 available assessments[c]	Avg. FU	8.19	10.48	8.27	7.49	2.48	5.74	7.95	3.48	7.17
		# Obs.	265	15	280	153	33	186	418	48	466
		# Subj.	25	1	26	13	7	20	38	8	46
	≥2 available assessments[c]	Avg. FU	8.19	10.48	8.27	8.12	4.34	7.17	8.16	5.57	7.85
		# Obs.	265	15	280	152	30	182	417	45	462
		# Subj.	25	1	26	12	4	16	37	5	42
BMI	≥1 available assessments	Avg. FU	9.08	8.89	9.07	7.04	4.63	6.24	8.49	5.48	8.02
		# Obs.	562	33	595	201	75	276	763	108	871
		# Subj.	39	2	41	16	8	24	55	10	65
	≥2 available assessments	Avg. FU	9.08	8.89	9.07	7.04	5.29	6.51	8.49	6.09	8.15
		# Obs.	562	33	595	201	74	275	763	107	870
		# Subj.	39	2	41	16	7	23	55	9	64
TC	≥1 available assessments	Avg. FU	8.15	8.88	8.18	5.64	3.39	4.85	7.53	4.61	7.11
		# Obs.	481	24	505	124	36	160	605	60	665
		# Subj.	40	2	42	13	7	20	53	9	62
	≥2 available assessments	Avg. FU	8.15	8.88	8.18	6.11	5.93	6.07	7.68	6.91	7.60
		# Obs.	481	24	505	123	33	156	604	57	661
		# Subj.	40	2	42	12	4	16	52	6	58
HDLC	≥1 available assessments	Avg. FU	6.23	8.30	6.28	5.50	2.34	4.33	6.05	3.09	5.63
		# Obs.	302	8	310	114	26	140	416	34	450
		# Subj.	37	1	38	12	7	19	49	8	57
	≥2 available assessments	Avg. FU	6.58	8.30	6.63	5.99	5.46	5.88	6.44	6.17	6.42
		# Obs.	300	8	308	113	22	135	413	30	443
		# Subj.	35	1	36	11	3	14	46	4	50

SBP: Systolic blood pressure. QTc: Corrected QT interval duration. BMI: Body mass index. TC: Total cholesterol. HDLC: High-density lipoprotein cholesterol. Glu: Fasting glucose. SCORE: Systematic Coronary Risk Evaluation. FU: Follow-up in years. [a]“East”/“West”: Czech Republic, Poland/Germany, Austria. [b]“Younger”/“Older”: Aged <60/≥60 yrs. at last available examination. [c]All pilots with at least one/two available assessment(s) of respective outcome.

Supplementary Table SI. Number of subjects, number of observations, and average follow-up duration by region of Europe, and pilot age

			East[a]			**West[a]**			**Total**		
			Younger[b]	**Older[b]**	**Total**	**Younger**	**Older**	**Total**	**Younger**	**Older**	**Total**
Glu	≥1 available assessments[c]	Avg. FU	7.76	8.88	7.82	3.86	3.09	3.57	6.81	4.25	6.40
		# Obs.	419	24	443	59	46	105	478	70	548
		# Subj.	40	2	42	13	8	21	53	10	63
	≥2 available assessments[c]	Avg. FU	7.76	8.88	7.82	4.57	6.17	5.00	7.07	7.08	7.07
		# Obs.	419	24	443	57	42	99	476	66	542
		# Subj.	40	2	42	11	4	15	51	6	57
SCORE	≥1 available assessments	Avg. FU	8.12	8.88	8.16	5.49	1.97	4.26	7.48	3.50	6.90
		# Obs.	477	24	501	118	26	144	595	50	645
		# Subj.	40	2	42	13	7	20	53	9	62
	≥2 available assessments	Avg. FU	8.12	8.88	8.16	6.49	3.44	5.68	7.77	5.26	7.51
		# Obs.	477	24	501	116	23	139	593	47	640
		# Subj.	40	2	42	11	4	15	51	6	57

SBP: Systolic blood pressure. QTc: Corrected QT interval duration. BMI: Body mass index. TC: Total cholesterol. HDLC: High-density lipoprotein cholesterol. Glu: Fasting glucose. SCORE: Systematic Coronary Risk Evaluation. FU: Follow-up in years. [a]“East”/“West”: Czech Republic, Poland/Germany, Austria. [b]“Younger”/“Older”: Aged <60/≥60 yrs. at last available examination. [c]All pilots with at least one/two available assessment(s) of respective outcome.

Supplementary Table SII. Linear mixed model coefficients for cardiometabolic risk factors and fatal cardiovascular disease risk prediction in HEMS pilots, including current smoking status as covariate.

		SBP [mmHg]		**QTc [ms]**		**BMI [kg/m^2]**		**TC [mmol/l]**		**HDLC [mmol/l]**		**Glu [mmol/l]**		**SCORE (6 m.) [%][a,b]**	
		Coef.	**95% CI**	**Coef.**	**95% CI**	**Coef.**	**95% CI**	**Coef.**	**95% CI**	**Coef.**	**95% CI**	**Coef.**	**95% CI**	**Coef.**	**95% CI**
Intercept[c]		126.6	122.7, 130.6	402.3	393.1, 411.4	25.3	23.7, 26.9	5.30	4.87, 5.72	1.24	1.08, 1.41	4.70	4.45, 4.95	0.004	0.003, 0.005
East European[d]		3.62	-0.94, 8.178	-10.31	-21.56, 0.95	1.33	-0.59, 3.24	-0.03	-0.51, 0.44	0.21[h]	0.03, 0.39	0.46[h]	0.18, 0.74	2.27[h]	1.80, 2.87
Smoking		-0.54	-3.15, 2.08	-0.63	-8.39, 7.14	-0.25	-0.52, 0.03	0.16	-0.09, 0.40	-0.07	-0.16, 0.02	-0.05	-0.23, 0.14	2.13[h]	1.91, 2.38
Relevant medication[e]		-0.74	-3.98, 2.50	---	---	---	---	-1.34[h]	-1.59, -1.08	-0.07	-0.17, 0.04	---	---	0.79[h]	0.72, 0.86
Age at first available assessment [yrs.][f]		0.02	-0.24, 0.28	0.36	-0.26, 0.98	0.09	-0.02, 0.20	0.01	-0.01, 0.04	0.02[h]	0.01, 0.03	0.02[h]	0.00, 0.03	1.18[h]	1.17, 1.20
Time since first available assessment [yrs.]	Younger[g]	0.05	-0.14, 0.24	1.22[h]	0.74, 1.69	0.12[h]	0.11, 0.14	-0.01	-0.02, 0.01	-0.01	-0.01, 0.00	0.02[h]	0.00, 0.03	1.17[h]	1.16, 1.17
	Older[g]	0.19	-0.40, 0.78	1.33	-0.50, 3.15	0.02	-0.04, 0.08	-0.07[h]	-0.12, -0.01	0.03[h]	0.01, 0.06	-0.03	-0.07, 0.01	1.10[h]	1.07, 1.13
	Difference[g]	0.14	-0.47, 0.75	0.11	-1.76, 1.97	-0.11[h]	-0.17, -0.04	-0.06[h]	-0.12, -0.01	0.04[h]	0.01, 0.07	-0.05[h]	-0.09, -0.01	0.94[h]	0.91, 0.97

SBP: Systolic blood pressure. QTc: Corrected QT interval duration. BMI: Body mass index. TC: Total cholesterol. HDLC: High-density lipoprotein cholesterol. Glu: Fasting glucose. SCORE: Systematic Coronary Risk Evaluation. CI: Confidence interval. [a]Probability to suffer a fatal cardiovascular event within 6 months as estimated by the SCORE algorithm. [b]Coefficients are multiplicative for this outcome (vs. additive for all others). [c]40-yr. old West European not currently using relevant medication, at first available assessment. [d]Pilot from Czech Republic or Poland (vs. Germany or Austria). [e]SBP: Blood-pressure lowering drugs, TC, HDLC: Lipid-lowering drugs, SCORE: Either type of drug. [f]Centered at 40 years. [g]"Younger"/"Older": Time effect for pilots aged <60/≥60 yrs. at last available examination, "Difference": Difference between these effects. [h]95% CI for covariate effect does not include null value (1 for SCORE, 0 for all other outcomes).

Supplementary Table SIII. Linear mixed model regression coefficients for cardiometabolic risk factors and fatal cardiovascular disease risk prediction in HEMS pilots, stratified by region of origin.

			SBP [mmHg]	**QTc [ms]**	**BMI [kg/m^2]**	**TC [mmol/l]**	**HDLC [mmol/l]**	**Glu [mmol/l]**	**SCORE (6 m.) [%][a]**
Intercept[b]		East[c]	129.9	390.4	26.6	5.32	1.45	5.12	0.011
		West[c]	127.1	410.8	25.6	5.71	1.17	4.94	0.007
Relevant medication[d]		East	-1.456	---	---	-1.452	-0.095	---	0.663
		West	1.913	---	---	-0.862	0.037	---	1.014
Age at first available assessment [yrs.][e]		East	0.020	0.655	0.095	0.034	0.012	0.019	1.214
		West	-0.045	-0.289	0.060	-0.038	0.022	0.004	1.131
Time since first available assessment [yrs.]	Younger[f]	East	0.129	1.595	0.112	-0.012	-0.004	0.022	1.181
		West	-0.248	0.478	0.153	0.018	-0.010	-0.024	1.136
	Older[f]	East	-0.278	1.480	0.134	-0.106	0.010	0.012	1.068
		West	0.774	1.007	-0.049	-0.021	0.058	-0.083	1.124
	Difference[f]	East	-0.406	-0.115	0.022	-0.094	0.014	-0.011	0.904
		West	1.022	0.529	-0.202	-0.039	0.068	-0.059	0.990

Note: Confidence intervals not included because of very small sample sizes (compare Web Appendix Table 1). SBP: Systolic blood pressure. QTc: Corrected QT interval duration. BMI: Body mass index. TC: Total cholesterol. HDLC: High-density lipoprotein cholesterol. Glu: Fasting glucose. SCORE: Systematic Coronary Risk Evaluation. CI: Confidence interval. AIC: Akaike’s Information Criterion. ICC: Intraclass coefficient. [a]Probability to suffer a fatal cardiovascular event within 6 months as estimated by the SCORE algorithm. Coefficients are multiplicative for this outcome (vs. additive for all others). [b]40-yr. old not currently using any relevant medication, at first available assessment. [c]“East”/“West”: Czech Republic, Poland/Germany, Austria. [d]SBP: Blood-pressure lowering drugs, TC, HDLC: Lipid-lowering drugs, SCORE: Either type of drug. [e]Centered at 40 years. [f]“Younger”/“Older”: Time effect for pilots aged <60/≥60 yrs. at last available examination, “Difference”: Difference between these effects.

3. Publication 2: Helicopter simulator performance prediction

Bauer, H., Nowak, D., & Herbig, B. (2018). Helicopter simulator performance prediction using the Random Forest method. *Aerospace Medicine and Human Performance*, *89, 967–975*. https://doi.org/10.3357/AMHP.5086.2018

Helicopter Simulator Performance Prediction Using the Random Forest Method

Hans Bauer, M.Sc.[a,*], Dennis Nowak, M.D.[a], and Britta Herbig, Ph.D.[a]

[a]Institute and Clinic for Occupational, Social, and Environmental Medicine, University Hospital of LMU Munich, Munich, Germany.

[*]Corresponding author. Postal address: Institut und Poliklinik für Arbeits-, Sozial- und Umweltmedizin, Ziemssenstraße 1, 80336 München, GERMANY. Phone: +49 89 4400 55305, Fax: +49 89 4400 54564, Email: hans.bauer@med.uni-muenchen.de.

Running Head: PERFORMANCE & RANDOM FORESTS

Abstract Word Count: 211

Main Text Word Count: 5616

Number of References: 29

Number of Tables: 2

Number of Figures: 2

Abstract

INTRODUCTION: Different aspects of the aviation system, such as pilot's fitness, supervision, and working conditions, interact to produce or protect against flight safety hazards. Machine learning methods such as Random Forests may help identify system characteristics with the potential to affect flight safety from the large number of candidate predictors that results when multiple system levels are considered simultaneously. METHODS: Fifty-four pilot-related and occupational candidate predictors of simulator flight performance in two malfunction scenarios completed by 51 male European helicopter emergency medical services pilots were derived from pilots' self-report questionnaires and aeromedical examination records. In a cross-sectional explorative analysis, the Random Forest method was used to screen for informative predictors. Predictors scoring above the critical threshold for the conditional permutation variable importance (VI) statistic were selected. RESULTS: In five predictors, the VI statistic averaged across 2,000 Random Forest runs exceeded the selection threshold: Higher perceived rewards (VI=0.0691) and predictability (VI=0.0501) at work were associated with higher performance scores, and higher physiological dysregulation (VI=0.0495) and alanine aminotransferase (VI=0.0224) with lower scores. Performance also differed between the simulators at the two training sites (VI=0.0298). DISCUSSION: Random Forests may usefully complement previously applied methods for the identification of human factors safety hazards. The identified performance predictors suggest further areas with potential for safety improvements.

Keywords: Flight safety; Machine Learning; Helicopter pilots; Human factors; Flight simulator

Introduction

Improving aviation safety has become an increasingly challenging task since most easily recognizable hazards, such as technological deficiencies, have been reduced significantly over the last decades and the remaining ones are often of a latent or insidious nature. Current human factors models of accident causation such as the popular “Swiss cheese” model[16] therefore take a systemic perspective acknowledging hazards related to both the aircraft operators themselves and their organizational environment.

Nevertheless, in empirical studies of aircraft accident causes based on such a systemic perspective, there is usually a notable gradient in the frequency of identified contributory factors across the system levels, with unsafe operator acts being identified most often and organizational influences least often[17]. It seems likely that the absence of organizational, supervisory, and working condition-related factors is at least in part due to the reliance on accident investigation board reports, which vary in scope and may investigate more distal conditions surrounding an accident only in the most severe cases[27]. Further complicating the issue, information on the relevant comparison group (i.e., pilots that did not experience an accident or incident) is usually not available or only in aggregated form in report-based retrospective studies.

These issues can be addressed through prospective studies, which however introduce different methodological problems. Aviation safety-related prospective studies can only use proxy outcomes such as line check ratings or simulator performance, and are often based on small samples, which is problematic given the large number of potentially interacting hazardous and protective factors operating at different system levels. Isolating the independent contribution of individual system characteristics to risk is difficult when samples are small relative to the number of characteristics; conventional techniques such as univariate pre-screening of candidate predictors or multiple regression-based stepwise variable selection are problematic since they do not consider the effects of discarded predictors and may result in upwardly biased effect estimates[7]. They furthermore generally assume linear additive relationships between predictors and outcome.

Several alternative procedures have been proposed to solve the problem of selecting influential predictors[28]. One approach which has recently gained popularity is based on Random Forests, a machine learning technique which is appropriate for situations when the number of predictors is large relative to the sample size, avoids overfitting, and automatically takes into account interactions between predictors, while also generating a measure of predictor importance which can be used for variable selection[23]. It is therefore well-suited for application to the aviation hazard identification problem outlined above, and its capability for predictor identification has, to the best of our knowledge, so far not yet been utilized in the aviation safety field.

In this cross-sectional explorative study, we therefore apply the Random Forest method in order to identify potential aviation safety hazards by selecting the most powerful predictors of simulator flight performance in a sample of professional helicopter pilots from a deliberately broad set of predictors covering both personal and occupational human factors safety aspects. The predictors are derived from pilot questionnaire self-reports and aeromedical fitness examination records, and performance is assessed through flight instructor ratings of the pilots' handling of two naturalistic system malfunction scenarios. Our aim is to evaluate, in an aviation safety context, whether the Random Forest method is producing meaningfully interpretable results when the number of predictors is large relative to the sample size.

Methods

Subjects

The analysis reported herein is based on data from a study on age and flight safety in helicopter emergency medical services (HEMS). Active HEMS pilots employed by one of five air rescue operators (two based in Germany and one each in Austria, Poland, and the Czech Republic) who completed a simulator flight during the data collection period from September 2015 to October 2016 at training sites in Warsaw, Poland (Polish operator) or in Hangelar, Germany (all other operators) were asked to participate. Although study participation was open to both genders, the operators' HEMS pilot workforces consisted almost exclusively of men at the time of recruitment, resulting in a male-only sample. The study had been approved by the Ethics Committee at Munich University's Faculty of Medicine (Project No. 466-15), and written informed consent had been obtained from all study subjects, prior to data collection. Subjects were able to separately indicate their consent to the collection of simulator performance data and of aeromedical examination record data.

Materials and Procedure

We collected data from the three distinct sources: (1) standardized ratings of simulator flight performance in two malfunction scenarios made by training instructors, (2) self-report questionnaires from participating pilots, and (3) records of the participating pilots' statutory aeromedical examinations.

Subjects completed a simulator flight for training or testing purposes as mandated by their employer or the responsible regulatory authority in order to maintain currency of their helicopter type rating or of operating procedures, including emergency procedures. Each session consisted of a pre-flight briefing of the pilot by the flight instructor, the actual simulator flight, and a short post-flight debriefing. At Hangelar, the flights were conducted in motion-capable full flight simulators. A non-motion-capable flight training device was used in

Warsaw. At both sites, the simulators corresponded to the helicopter types flown by the pilots during their actual duty and could simulate different geographies including urban structures and weather conditions.

Embedded in this regular training/check flight, which was independent of the study, the responsible flight instructor – who had previously been briefed by a member of the research team – deployed two study malfunction scenarios, which together took about ten minutes to complete, and rated the pilot's responses according to a standardized rating sheet. Several experienced HEMS pilots had been consulted prior to data collection in order to select relevant scenarios and to develop the corresponding rating scheme.

The first scenario, "Transmission oil system malfunction", was designed to assess the pilot's vigilance and situational awareness. It involved the timely detection of a gauge indicating helicopter transmission oil status moving slowly towards a critical value, and terminated as soon as the pilot detected the problem or the critical value was reached. More specifically, the malfunction concerned a gradual decrease in transmission oil pressure in the Hangelar simulators, and a gradual increase in transmission oil temperature in the Warsaw simulator, since the oil pressure decrease scenario could not be implemented in the latter simulator. The respective gauges were located next to each other at similar positions within the flight instrument panel in all simulators, and the corresponding malfunctions are comparable in terms of their implications for flight safety. We therefore assumed the tasks to have similar properties in terms of their demands to situational awareness. However, the time from initiation of malfunction to hitting the critical value, which was defined as the first occurrence of an additional optical or acoustic warning signal by the system, was approximately twice as long in the Hangelar simulator, compared to the Warsaw simulator (170 vs. 84 sec). Recognition of the malfunction by the pilot ahead of the critical value was rewarded with two points (same weight as an “important” sub-task of the second scenario described below).

The “Tail rotor drive failure” scenario constituted a complex emergency situation which required the pilot to bring the aircraft safely to the ground via a so-called autorotation procedure. Instructors rated performance of nine subtasks which involved situational awareness, decision-making, knowledge of procedures, spatial orientation, and psychomotor control, for a maximum of twelve points. Proper completion of subtasks yielded one or two points, based on their importance to the successful completion of the entire procedure.

Fifteen simulator flights were concurrently rated by two instructors, with good interrater agreement over a total of 141 binary decisions (subtask ratings), Cohen’s $\kappa = .91$ (94.3% concordant decisions).

In the briefing parts of the session, pilots were asked to fill in short questionnaires inquiring about flight experience and current self-rated health (pre-flight briefing), and about subjectively experienced strain and simulator sickness during the flight as well as risk-seeking propensity (post-flight debriefing). Additionally, the pilots were handed over a longer questionnaire covering working conditions, subjective experiences at work, and general sense of well-being, which they were asked to fill in and send back over the course of the next few days. Wherever possible, questionnaire scales and items had been taken from published, psychometrically validated instruments such as the Copenhagen Psychosocial Questionnaire. They were included based on their relevance regarding health, work performance and safety[14,20].

Consenting pilots were also asked to provide details of the aeromedical examiners and centers they had visited during the preceding ten years, and release them from their non-disclosure duty. We mailed requests for transfer of full aeromedical examination documents to all specified physicians/centers, and sent a reminder letter after four weeks. If the reminder also did not elicit a response, we made at least one more contact attempt via telephone or email. We received digitalized or paper documents dated between April 2004 and July 2016 from 23 physicians/centers (61% of 38 contacted). All documents were searched for information concerning quantitative clinical measurements of any kind, as well as smoking and medication. Entry of the corresponding data was conducted according to a coding manual developed after an initial review of the received documents and continuously revised during the data entry process.

Statistical Analysis

Our main outcome variable, simulator performance, was calculated as the sum of instructor ratings for the two study scenarios (ranging theoretically between a minimum of 0 and a maximum of 14 attainable points). The predictor variables were based on the self-report questionnaires and aeromedical examination findings. Questionnaire-based variables were calculated as scale means or sum scores, or (in the case of single-item measures) the untransformed item scores. We selected individual medical risk markers for analysis based on their availability in the aeromedical examination findings data and on their association with health status and incapacitation risk of a pilot. Moreover, we included as predictors two composite indices based on the individual risk markers: Risk of a fatal cardiovascular event within six months according to the SCORE algorithm[5], and a "physiological dysregulation index" based on the gerontological concept of "biological aging"[4], which can be viewed as a subclinical trajectory towards frailty and disease due to insidious functional decline in multiple organ systems[13]. Both cardiovascular event risk scores such as SCORE and physiological dysregulation indices have been found to be associated with cognitive decline[2,9].

Our dysregulation index is based on all available measurements of 18 health risk-related biomarkers of cardiovascular, metabolic, liver, kidney, immune, hematologic, and ocular function as well as hearing level. For each biomarker, a subject's probability to have a biomarker reading within an "unhealthy" range (e.g., systolic blood pressure >140 mmHg) was estimated based on the longitudinal individual-specific distribution of biomarker readings using a linear mixed model. The dysregulation score is the sum of these probabilities (thus ranging theoretically between 0 and 18) and can be understood as a pilot's expected number of biomarker readings outside the healthy range (see Auxiliary Appendix A for methodological details).

Finally, we also included simulator type as a predictor because of the aforementioned differences in the transmission oil failure scenario and in motion capability. Only variables where less than one third of subjects had missing data were used for the analysis. In total, this resulted in a set of 54 predictors which can be categorized as psychosocial and physical work stressors (including protective factors such as social support), psychosocial and physical strain symptoms, other aspects of working conditions (e.g., working hours), medical risk markers, subjective experience of the pilot during the simulator training session, and general pilot characteristics (Table I; see Auxiliary Appendix B for further details). For all medical risk markers except physiological dysregulation (which was based on the complete longitudinal information available), we used only the latest available assessment in the simulator performance prediction.

Table I also shows the number of missing data points per predictor. To impute missing values, we used the R implementation of the missForest algorithm[21], which iteratively applies the Random Forest method (described below). In order to account for uncertainty in the imputation estimate, we created a total of 20 imputed datasets and compared or aggregated analysis results across these datasets where appropriate.

After the imputation step, we applied Random Forests for simulator performance prediction and variable selection. Random Forest is a supervised machine learning method consisting of an ensemble of decision trees which attempt to predict the outcome by defining, based on predictor values, groups that are homogeneous with respect to the outcome. In Random Forests, many such trees are “grown” on random sub-samples of the study subjects and predictor variables, and the single trees’ predictions are averaged to produce the Forest’s prediction. Since each component tree is trained only on a random subsample, the remaining (“out-of-bag”) cases can be conveniently used as an “external” validation set for the prediction quality of this particular tree. Prediction errors for the out-of-bag cases can again be averaged across all trees to yield a measure of prediction accuracy which is less affected by overfitting. Furthermore, and particularly important in the present context, Random Forests are also able to produce a measure of relative importance in the prediction of the outcome for each individual predictor variable even when the number of predictors is larger than the sample size (which is not possible with conventional regression modelling techniques). In this way, relevant predictors can be identified. The importance of a predictor is calculated as the difference in out-of-bag prediction accuracy between a Random Forest grown on the original input data and that of another forest which is identical in parameterization and input data except that the values of the predictor have been randomly permuted, reducing its predictive capacity to chance level[23]. We used a modification of this variable importance measure (termed “conditional permutation importance”) which accounts for intercorrelations among predictors (analogous to the mutual adjustment of covariate effects in a multiple regression)[22].

For each of the 20 imputed datasets, we fitted 100 Random Forests and calculated the mean variable importance across the resultant 2,000 Forests for each predictor variable, to obtain an estimate which is less affected by the random variation inherent in the Random Forest procedure. We then selected those variables for further inspection whose mean variable importance exceeded the absolute value of the minimum mean variable importance among all variables. This selection criterion was suggested as an improvement over the z-score metric commonly used for Random Forest variable selection[23]. Note that conventionally reported statistics such as p values or confidence intervals are not directly applicable to Random Forests, although the described selection criterion can be considered an analogue to a p-value-based statistical significance threshold such as $p < .05$.

We also examined the stability of the importance assigned to a variable by inspecting how strongly the variable's position in the importance ranking varied between the 2,000 Random Forest fits. To visualize the relationship between the selected predictors and the outcome, we present partial dependence plots[10], which display the effect of a predictor across its value range averaged over all other predictor value combinations occurring in the sample (i.e., the estimated marginal effect of the predictor).

All statistical analyses reported subsequently were done using the R statistical software, version 3.3.1.

Results

Figure 1 shows the study subject flow. Unavailability of simulator data was mostly due to a mismatch of pilots' training session schedules with the data collection period, whereas nonresponse by physicians/centers from whom examination records had been requested was the main reason for unavailability of aeromedical data. The analysis sample consisted of 51 male pilots with all data types available, 15 (29%) from the Western European countries (Germany, Austria) and 36 (71%) from the Eastern European countries (Poland, Czech Republic); see Table I for descriptives. Simulator performance scores were concentrated at the upper end of the scale and slightly left-skewed (range: 7–14, median: 12, skewness: -0.39). For some of the medical predictors, there was a considerable time lag between their assessment and the simulator training session.

Mean variable importances of the 54 predictor variables ranged from -0.0196 to 0.0691. Five variables had a higher mean importance score than the selection threshold of

0.0196 (see Section “Statistical Analysis”): The reward subscale of the Effort-Reward-Imbalance inventory[18] measuring perceived rewards at work, the predictability of work demands score[26], the physiological dysregulation score, simulator type, and the last available measurement of alanine amino-transferase, a marker of liver pathology (Table II; see Auxiliary Appendix C for details on all predictors). There is a notable drop in mean variable importance from the selected to the unselected variables. Perceived rewards clearly stands out as having the highest predictive power, followed by perceived predictability and dysregulation and, after another distinct drop in variable importance, by simulator type and alanine aminotransferase. The importance rankings were fairly stable across Random Forest fits for the reward, predictability, and physiological dysregulation variables. Rankings were more unstable for simulator type and in particular for alanine aminotransferase.

The estimated marginal effects of the selected variables are illustrated in Figure 2. Even in those variables which have the strongest association to simulator performance in the sample, the effects on performance are all quite small and on the order of 0.1 to 0.2 performance score points (where one performance score point roughly corresponds to one mistake in the simulator scenarios). In other words, simulator performance is, on the whole, poorly predicted by the set of 54 variables considered. Still, the direction of effects is generally plausible in the selected variables: Performance was better in those who perceived their work as more rewarding and work demands as more predictable, and worse in those with higher physiological dysregulation scores and alanine aminotransferase levels. Several possible explanations come to mind (e.g., differences in simulator handling characteristics, in the transmission oil malfunction scenario, or in the response tendencies of the involved instructors) regarding the effect of simulator type. However, since this variable was included for technical reasons (i.e., adjustment for simulator idiosyncrasies), rather than theoretical reasons, we will not discuss it any further.

The predicted effects in the quantitative variables were not linear but rather stepwise. Note that for the working conditions variables, the threshold values (~12 for the Reward Subscale and ~3 for the Predictability scale) correspond to the theoretical scale means, that is, performance was predicted to be better in those who, on the whole, agreed to statements such as “I receive the respect I deserve from my superiors”, and/or indicated that all in all, daily job demands could be predicted to a large extent. The threshold effect in the physiological dysregulation score can be interpreted such that pilots who are in an excellent overall state of health according to the biomarkers used for the dysregulation score (i.e., all biomarker levels are well within the healthy range) are predicted to perform better than those who tend to have levels near or beyond the limits of the healthy range in at least one of the biomarkers. Although the alanine aminotransferase effect is harder to interpret, it is noteworthy that the threshold value after which performance decreases (~40 U/l) falls into the upper end of male population reference range limits (e.g., [15]).

Discussion

In this cross-sectional explorative study, we applied the Random Forest machine learning method for the selection of the most influential predictors of HEMS pilot performance in two simulated in-flight failure scenarios from a set of predictors which covered personal and occupational human factors aspects potentially relevant to flight safety and which was larger than the sample size. Although the predictors on the whole explained rather little of the variation in simulator flight performance, five of them (perceived rewards at work, perceived predictability of work demands, physiological dysregulation, alanine aminotransferase, and simulator type) explained more than would be expected by chance alone. Their effects appeared to be stepwise rather than linear, and their direction was mostly consistent with theoretical expectations.

To the best of our knowledge, this is the first application of the Random Forest method to identify potential human factors safety issues. Many analyses use conventional bivariate or multiple regression methods[3,25]. These “classical” methods have good statistical properties when their assumptions are met (which notably includes the rather restrictive assumption of linear effects), and are well-suited for confirmatory analyses investigating the effects of a smaller number of predictors of interest. In contrast, machine learning methods appear to be better suited for explorative analyses where vast amounts of information on potential predictors is available and little is known about the functional relationship between the predictors and the outcome.

Analysis of natural-language documents are a prototypical example of a high-dimensional input problems; in aviation safety, machine learning methods are becoming increasingly popular for text mining of accident/incident report narratives. Often, unsupervised learning methods are applied to cluster occurrence reports and subsequently identify common underlying themes, such as cigarette smoking by passengers[24]. While these approaches are very flexible and can accommodate data which is otherwise difficult to process, they are able to identify only those factors mentioned in the reports, which tend to be proximal factors[27]. Furthermore, human factors hazards such as “confusion” are often hard to isolate for text mining algorithms since their description mostly lacks highly distinctive signaling words which characterize the more technical issues[24]. Our analysis may be located somewhere in between the two extremes of linear modelling of selected features and indiscriminate text mining, in that it allows a certain preselection of features, including those which are usually not considered in occurrence reports such as organizational stressors, but does not impose restrictive assumptions on the relationship between predictors and outcome.

Of the five selected informative predictors, two measured aspects of psychosocial work stress. This kind of stress is known to affect safety at work[14]. Young[29] reviewed the effects of life stress (including work stress) on pilot performance and suggested that life stress might undermine performance by increasing fatigue (through reduced sleep quantity and quality as well as emotional exhaustion), decreasing motivation to perform (e.g., skipping "unimportant" tasks such as checklist procedures), worsening interpersonal relations and communication with colleagues, and increasing intrusive and distractive worrying. With regards to the most predictive of our variables, perceived rewards at work[18], reduced motivation may be a plausible contributing factor[11]. On the other hand, the second selected work stressor[26] assesses the degree to which the pilot perceives his work environment to be predictable. This may be related to instances of disrupted action regulation ("hindrance stressors") during work; repeated experience of disrupted action regulation might lead to a more passive style of coping with work demands[12].

The concept of physiological dysregulation as an overall loss of an organism's capacity to maintain homeostasis has recently received increased interest in gerontology to explain differences in the "healthiness" of aging between individuals[2,13]. Physiological dysregulation was found to be associated with cognitive decline and reduced psychomotor performance already at age 38[2], but analyses of the relation between dysregulation and work performance or safety are lacking so far. Especially in safety-critical jobs such as piloting, the use of physiological dysregulation indices appears to be an interesting concept for early detection of health-related risks at a subclinical stage[19]. Given the existing framework of aeromedical examinations in professional pilots, more systematic investigations of the effects of physiological dysregulation on flight performance might be implemented with comparatively little effort.

The second selected medical predictor, alanine aminotransferase, is a marker related to liver cell necrosis used in the diagnosis of liver conditions including alcoholic or nonalcoholic fatty liver disease. This result may evoke associations of a possible role of alcohol use[6], but clearly the exploratory nature of our findings, especially regarding this predictor which was quite unstable across Random Forest fits in terms of predictive power, prohibits any such speculations given the sensitive nature of the topic; however, for purposes of replication and confirmation, this marker might be included in future investigations of the relation between pilot health and performance.

Among the limitations to this study, the most immediately apparent is the small sample size, which highlights a drawback of our approach compared to, for example, the use of occurrence reports: Collecting a range of data sources, each with its own mechanisms of sample attrition, from an inherently small population (HEMS pilots) will almost inevitably lead to a small sample size. On the other hand, it should be kept in mind that our approach was motivated precisely by the question of whether it is possible to obtain interpretable results when the number of potential predictors is large relative to the sample size, a situation which is not uncommon in human factors aviation safety studies that are not purely based on retrospective review of occurrence reports or administrative records.

Moreover, as is the case with explorative research in general, there is also a threat of false-positive findings due to many simultaneously assessed associations. It should be noted, however, that the directions of the selected variables’ effects appear to be generally plausible and that there is a relatively clear separation between the selected variables and unselected variables in terms of the variable importance scores. The logic behind the chosen selection threshold also seems to imply some protection against capitalization on chance since the absolute value of the minimum observed variable importance score should be expected to increase with the number of noise predictor variables involved (whose variable importance should vary randomly around zero). Finally, in the light of the tradeoff between type I and type II errors in statistical decision-making, it has been suggested that there should be a focus on minimizing the latter error in aviation safety research due to the potentially grave implications of false-negative findings[8].

With regards to the aeromedical data, there was additionally the problem of a time lag between the last available assessment and the simulator session, which was very large (2-2.5 years) in some of the biomarkers, including alanine aminotransferase which had been selected as an informative predictor by the Random Forest procedure. According to a linear mixed model analyses of time-stability of biomarker levels we conducted earlier, between 44 and 95 percent of the total variation in biomarker levels across average follow-ups of 5.2 to 8.2 years were due to differences in pilot averages across time, indicating considerable stability of interindividual differences in the biomarkers (for alanine aminotransferase specifically, the respective figures were 70 percent and 5.2 years). One might therefore assume that differences at the time of last available assessment carried over to the simulator session to some extent.

Finally, in contrast to occurrence report studies, our outcome of simulator flight performance can be viewed only as a proxy to the actual outcome of interest. Thus, in order to achieve an optimal ecological validity, we consulted extensively with experienced HEMS pilots and flight instructors in the selection of malfunction scenarios as well as in devising the scoring procedure.

To conclude, we identified three well-interpretable predictors of HEMS pilot simulator flight performance (two occupational stressors and an index of physiological dysregulation) from a broad array of candidates by exploiting the capability of the Random Forest machine learning method to select important predictors even when their number is large relative to the sample size. The predictors were taken from different sources (self-report, medical examinations) and covered different aspects of potential relevance to the error chain as outlined in modern systemic human factors safety approaches. Although our study is explorative in nature which precludes confident statements about concrete measures to improve safety in HEMS, the results do suggest that the effect of working conditions and their perception by the pilots deserve further scrutiny. For example, the role of work stressors on HEMS pilots' subjective well-being might be investigated; well-being and mental health of professional pilots have received considerable attention recently[1], but data for HEMS pilots are lacking. The physiological dysregulation construct is an intriguing potential tool for early recognition of latent pathology in professional pilots. While our dysregulation index was of a somewhat ad-hoc nature constrained by data availability, the utility of current measures of dysregulation derived from theoretical considerations[2,13] for screening, prevention and selection purposes in aeromedical examinations might be further investigated. Finally, our study showcases the potential of the Random Forest method in the field of aviation human factors. For example, it could be applied to appropriately quantified information from accident investigation databases to identify factors associated with accident lethality. However, abundant data are also collected in everyday aviation operations, and with some effort invested into database normalization, information about the effect of operative conditions (e.g. timing of missions, weather, geographical location) on mission safety parameters may be quantitatively analyzed by individual HEMS operators using Random Forests. In a more ambitious approach, a framework for a common harmonized database which might include organizational, operational, administrative, and even aeromedical information could be established between operators or also between different aviation sectors. With such a large-scale database, the full potential of machine learning methods, which are designed to handle large amounts of information, could be brought to bear. Careful consideration would need to be given to feasibility (e.g. due to data comparability, data protection, and confidentiality issues) with this approach. In any case, the presented use of the Random Forest method may be a fruitful addition to existing risk analysis tools, helping operators to think “outside the box” in their efforts to identify additional flight safety measures.

Acknowledgments

Funding & Conflicts of Interest

This work was supported by an unrestricted educational grant from the European HEMS & Air Ambulance Committee e.V. (EHAC) to the Institute and Clinic for Occupational, Social, and Environmental Medicine. Freedom from interference in any stage of the research process was laid down contractually between EHAC and the research project team, represented by the manuscript authors. EHAC had no role in the formulation of the study design, the collection, analysis and interpretation of the data, the writing of the report, and the decision to submit the paper for publication.

References

1. Aerospace Medical Association. Pilot Mental Health: Expert Working Group Recommendations – Revised 2015. Aerosp Med Hum Perform 2016; 87(5):505–507.
2. Belsky DW, Caspi A, Houts R, Cohen HJ, Corcoran DL, Danese A, et al. Quantification of biological aging in young adults. Proc Natl Acad Sci U S A 2015; 112(30):E4104–E4110.
3. Boyd DD. Causes and risk factors for fatal accidents in non-commercial twin engine piston general aviation aircraft. Accid Anal Prev; 77:113–119.
4. Burkle A, Moreno-Villanueva M, Bernhard J, Blasco M, Zondag G, Hoeijmakers JH, et al. MARK-AGE biomarkers of ageing. Mech Ageing Dev 2015; 151:2–12.
5. Conroy RM, Pyörälä K, Fitzgerald AP, Sans S, Menotti A, de Backer G, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. Eur Heart J 2003; 24(11):987–1003.

6. Cook CC. Alcohol and aviation. Addiction 1997; 92(5):539–555.
7. Harrell FE. Regression modeling strategies. With applications to linear models, logistic regression, and survival analysis. 2nd ed. Cham: Springer; 2015: 67–72.
8. Harris D. The importance of the Type II error in aviation safety research. In: Farmer E, editor. Stress and error in aviation. Aldershot: Avebury Technical; 1991:151–157.
9. Harrison SL, Ding J, Tang EY, Siervo M, Robinson L, Jagger C, et al. Cardiovascular disease risk models and longitudinal changes in cognition: a systematic review. PLoS One 2014; 9(12):e114431.
10. Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning. Data mining, inference, and prediction. 2nd ed. New York, NY: Springer; 2017: 369–370.
11. Judge TA, Thoresen CJ, Bono JE, Patton GK. The job satisfaction-job performance relationship: a qualitative and quantitative review. Psychol Bull 2001; 127(3):376–407.
12. Lepine JA, Podsakoff NP, Lepine MA. A meta-analytic test of the challenge stressor - hindrance stressor framework. An explanation for inconsistent relationships among stressors and performance. Acad Manag J 2005; 48(5):764–775.
13. Milot E, Morissette-Thomas V, Li Q, Fried LP, Ferrucci L, Cohen AA. Trajectories of physiological dysregulation predicts mortality and health outcomes in a consistent manner across three populations. Mech Ageing Dev 2014; 141-142:56–63.
14. Nahrgang JD, Morgeson FP, Hofmann DA. Safety at work: a meta-analytic investigation of the link between job demands, job resources, burnout, engagement, and safety outcomes. J Appl Psychol 2011; 96(1):71–94.
15. Piton A, Poynard T, Imbert-Bismut F, Khalil L, Delattre J, Pelissier E, et al. Factors associated with serum alanine transaminase activity in healthy subjects: consequences for the definition of normal values, for selection of blood donors, and for patients with chronic hepatitis C. MULTIVIRC Group. Hepatology 1998; 27(5):1213–1219.

16. Reason J. Human error: models and management. BMJ 2000; 320(7237):768–770.

17. Shappell S, Detwiler C, Holcomb K, Hackworth C, Boquet A, Wiegmann DA. Human error and commercial aviation accidents: an analysis using the human factors analysis and classification system. Hum Factors 2007; 49(2):227–242.

18. Siegrist J, Dragano N, Nyberg ST, Lunau T, Alfredsson L, Erbel R, et al. Validating abbreviated measures of effort-reward imbalance at work in European cohort studies: the IPD-Work consortium. Int Arch Occup Environ Health 2014; 87(3):249–256.

19. Sluiter JK. High-demand jobs: age-related diversity in work ability? Appl Ergon 2006; 37(4):429–440.

20. Stansfeld S, Candy B. Psychosocial work environment and mental health – a meta-analytic review. Scand J Work Environ Health 2006; 32(6):443–462.

21. Stekhoven DJ, Bühlmann P. MissForest – non-parametric missing value imputation for mixed-type data. Bioinformatics 2012; 28(1):112–118.

22. Strobl C, Boulesteix A-L, Kneib T, Augustin T, Zeileis A. Conditional variable importance for random forests. BMC Bioinformatics 2008; 9:307.

23. Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychol Methods 2009; 14(4):323–348.

24. Tanguy L, Tulechki N, Urieli A, Hermann E, Raynal C. Natural language processing for aviation safety reports. From classification to interactive analysis. Comput Ind 2016; 78:80–95.

25. Taylor JL, O’Hara R, Mumenthaler MS, Yesavage JA. Relationship of CogScreen-AE to flight simulator performance and pilot age. Aviat Space Environ Med 2000; 71(4):373–380.

26. Tetrick LE, LaRocco JM. Understanding, prediction, and control as moderators of the relationships between perceived stress, satisfaction, and psychological well-being. J Appl Psychol 1987; 72(4):538–543.
27. von Thaden T, Wiegmann D, Shappell S. Organizational factors in commercial aviation accidents. Int J Aviat Psychol 2006; 16(3):239–261.
28. Walter S, Tiemeier H. Variable selection: current practice in epidemiological studies. Eur J Epidemiol 2009; 24(12):733–736.
29. Young JA. The effects of life-stress on pilot performance. Moffitt Field, CA: Ames Research Center; 2008.

Tables

Table I. Analysis variable sources and descriptives (mean / SD for quantitative variables, n / % for binary yes/no variables in italics)

Variable category	Variable	# of items / scale range	Source	# missing	Mean (SD) / *n (%)*	Assessment-simulator lag*
Outcome	Simulator performance	10 / 0–14	Rating by flight instructor	0	11.4 (2.1)	-
Psychosocial work stressors	Emotional demands	2 / 1–5	Pilot self-report (questionnaire)	0	2.3 (0.6)	-
	Social support	4 / 1–5	Pilot self-report (questionnaire)	0	3.4 (0.9)	-
	Work pace	2 / 1–5	Pilot self-report (questionnaire)	0	3.3 (0.7)	-
	Work predictability	3 / 1–5	Pilot self-report (questionnaire)	0	3.0 (0.8)	-
	Role clarity	3 / 1–5	Pilot self-report (questionnaire)	0	4.7 (0.4)	-
	Role conflict	4 / 1–5	Pilot self-report (questionnaire)	0	1.8 (0.6)	-
	Autonomy	3 / 1–5	Pilot self-report (questionnaire)	0	3.1 (0.7)	-
	Supervisor Feedback	3 / 1–5	Pilot self-report (questionnaire)	0	3.0 (1.1)	-
	Procedural Justice	4 / 1–5	Pilot self-report (questionnaire)	0	3.9 (0.8)	-
	Effort at work	2 / 2–8	Pilot self-report (questionnaire)	0	4.7 (1.6)	-
	Reward at work	5 / 5–20	Pilot self-report (questionnaire)	0	13.9 (2.2)	-
Physical work stressors	Physical demands, general	10 / 1–5	Pilot self-report (questionnaire)	0	3.2 (0.6)	-
	Physical demands, headgear	4 / 1–5	Pilot self-report (questionnaire)	0	2.5 (0.9)	-
Psychosocial strain	Irritation	6 / 1–7	Pilot self-report (questionnaire)	0	2.2 (0.7)	-
	Work engagement	9 / 0–6	Pilot self-report (questionnaire)	0	4.8 (0.8)	-
	Detachment from work	4 / 1–5	Pilot self-report (questionnaire)	0	3.2 (0.8)	-
	Subjective well-being	5 / 0–25	Pilot self-report (questionnaire)	0	19.0 (3.5)	-
	Energy / Fatigue	4 / 4–20	Pilot self-report (questionnaire)	0	17.2 (2.0)	-
Physical strain	# of body regions with pain†	13 / 0–13	Pilot self-report (questionnaire)	2	1.0 (1.7)	-
Other work-related factors	Work hours per month	-	Pilot self-report (questionnaire)	3	169 (27)	-
	Vacation days per year	-	Pilot self-report (questionnaire)	1	24.8 (7.7)	-
	Day shift duty?	-	*Pilot self-report (questionnaire)*	*2*	*48 (0.98)*	-
	Night shift duty?	-	*Pilot self-report (questionnaire)*	*2*	*32 (0.65)*	-
	24 hour stand-by shift duty?	-	*Pilot self-report (questionnaire)*	*4*	*5 (0.11)*	-
	Other shift type duty?	-	*Pilot self-report (questionnaire)*	*3*	*3 (0.06)*	-
	Any limit on flight time?	-	*Pilot self-report (questionnaire)*	*9*	*37 (0.88)*	-
Medical risk markers‡	*Smoking?*	-	*Aeromedical records*	*1*	*10 (0.20)*	*94*
	Any medication?	-	*Aeromedical records*	*1*	*9 (0.18)*	*89*
	Systolic blood pressure [mmHg]	-	Aeromedical records	0	130 (13)	92
	Resting heart rate [bpm]	-	Aeromedical records	0	67.8 (8.7)	92
	ECG QTc interval [ms]	-	Aeromedical records	15	406 (21)	91
	Body mass index [kg/m^2]	-	Aeromedical records	1	27.1 (3.3)	89
	Total cholesterol [mmol/l]	-	Aeromedical records	3	5.3 (0.8)	102
	HDL cholesterol [mmol/l]	-	Aeromedical records	8	1.5 (0.3)	349
	Triglycerides [mmol/l]	-	Aeromedical records	11	1.5 (0.7)	302
	Fasting glucose [mmol/l]	-	Aeromedical records	3	5.3 (0.6)	127
	Alanine aminotransferase [U/l]	-	Aeromedical records	14	37.1 (22.7)	712
	Aspartate aminotransferase [U/l]	-	Aeromedical records	14	28.2 (9.9)	712
	Serum creatinine [µmol/l]	-	Aeromedical records	8	91.6 (14.3)	886
	White blood cell count [10^3/µl]	-	Aeromedical records	2	6.5 (1.7)	101
	Hemoglobin [g/dl]	-	Aeromedical records	2	15.3 (1.0)	101
	Red blood cell distribution width [%]	-	Aeromedical records	4	13.0 (0.6)	101
	Intraocular pressure [mmHg]	-	Aeromedical records	9	15.2 (2.4)	213
	Hearing level at 3,000 Hz [dB HL]	-	Aeromedical records	4	16.6 (9.6)	268
	SCORE 6-month-risk [%]	-	Aeromedical records	3	0.07 (0.07)	102
	Physiological dysregulation§	-	Aeromedical records	¶	1.4 (0.9)	-
Experience during simulator training session	Self-rated health	1 / 1–5	Pilot self-report (questionnaire)	0	3.6 (0.9)	-
	Task load during flight	5 / 0–100	Pilot self-report (questionnaire)	0	49.0 (16.0)	-
	Simulator sickness	1 / 0–100	Pilot self-report (questionnaire)	0	14.4 (22.6)	-
General pilot characteristics	Age [years]	-	-	0	51.7 (8.2)	-
	# of real flight hours	-	Pilot self-report (questionnaire)	0	5340 (3513)	-
	# of simulator flight hours	-	Pilot self-report (questionnaire)	1	187 (411)	-
	Risk seeking	4 / 1–5	Pilot self-report (questionnaire)	0	1.4 (0.4)	-
Other	*Full flight simulator?***	-	-	*0*	*23 (0.45)*	-

Note. *Median time from risk marker variable assessment to simulator session in days. †Number of body regions (e.g., neck, lower back; 13 overall) where subject reported "occasional" or "frequent" pain. ‡As assessed at last available aeromedical examination (except physiological dysregulation). §Index derived from 18 health-risk associated biomarkers (see Auxiliary Appendix A). ¶Individual index component variables had been imputed before computation of index. Overall number of component variable missing values: 139 (=15.1% of 18*51 data points). **No = Flight training device.

Table II. Summary of variable importance characteristics of variables selected by Random Forest procedure (in italics) and of first three variables below selection criterion

	Mean variable importance	Variable importance rank order*			
		M	*SD*	Lowest	Highest
Reward subscale (ERI)	0.0691	1.42	0.72	6	1
Predictability of work demands scale	0.0501	2.55	1.17	16	1
Physiological dysregulation score	0.0495	2.61	1.18	14	1
Simulator type	0.0298	4.49	2.24	44	1
Last available ALT measurement	0.0224	6.40	5.19	54	1
Last available serum creatinine measurement	0.0113	9.66	6.03	47	2
Work hours per month	0.0099	10.66	6.79	48	2
Simulator sickness	0.0097	10.52	6.71	53	3

Note. Statistics calculated across 2,000 Random Forest fits (100 in each of 20 imputed datasets). Selection criterion: Mean variable importance greater than absolute value of minimum mean variable importance (-0.0196). ERI: Effort-Reward-Imbalance scale. ALT: Alanine aminotransferase. *Ranks (including mean ranks) coded such that lower values denote higher ranks (highest: 1st rank; lowest: 54th rank).

Figures

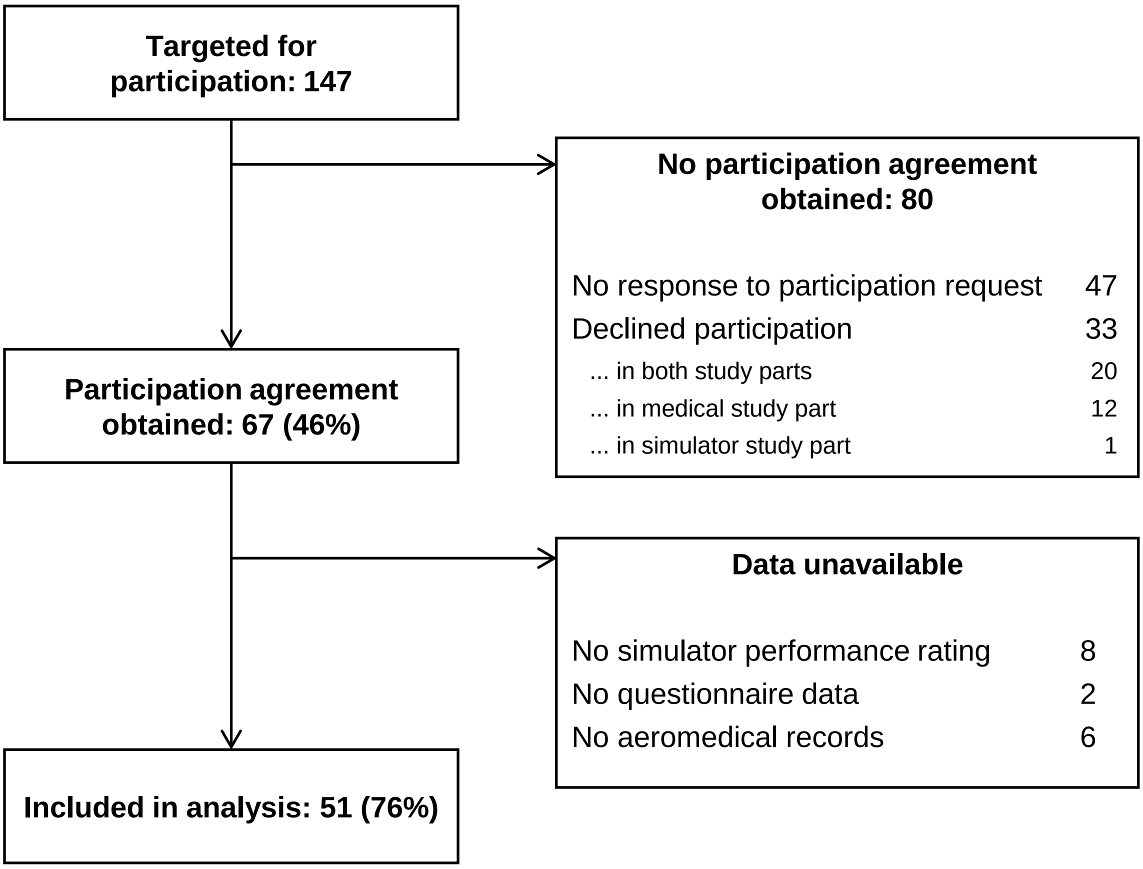


Figure 1. Study subject flow from recruitment to analysis stage.

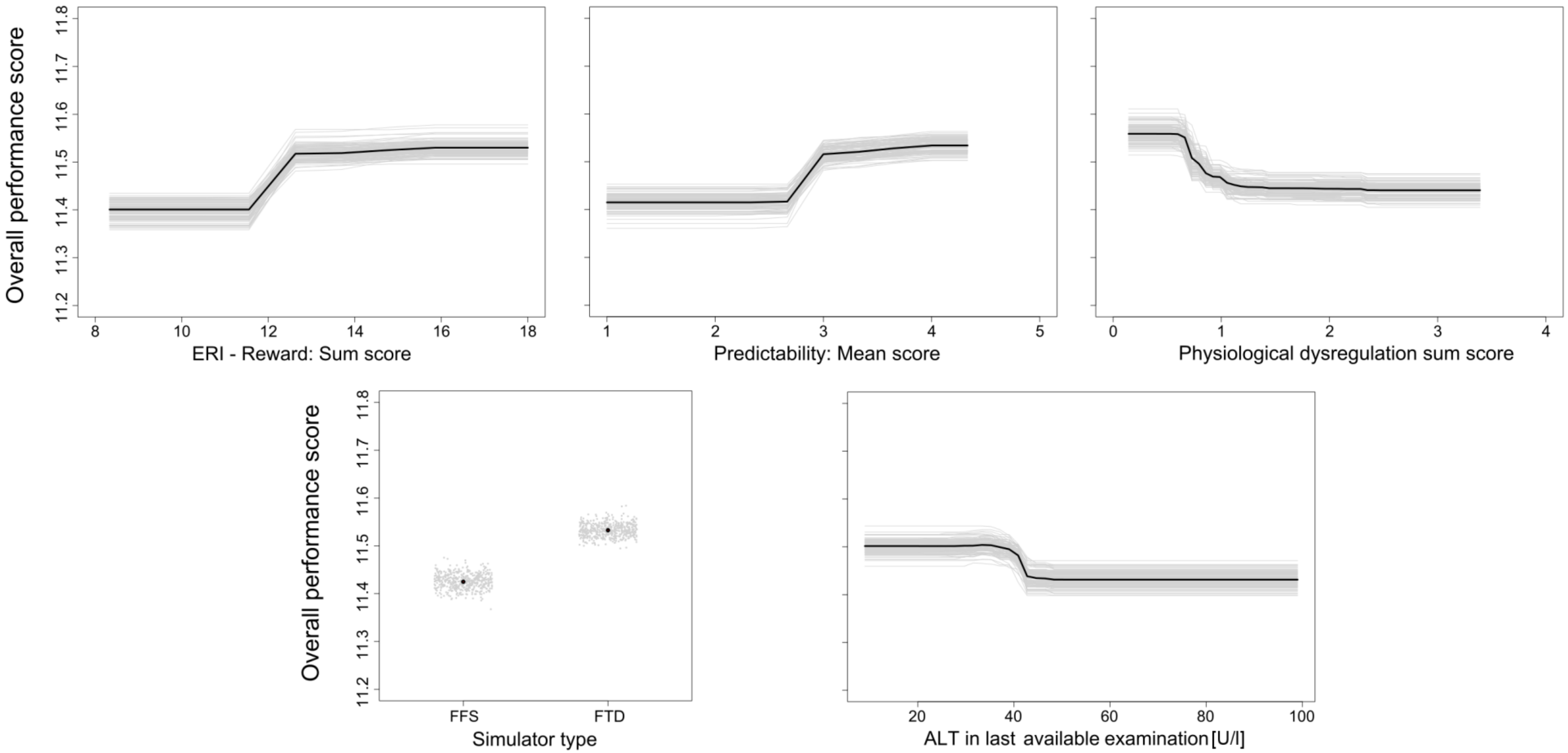


Figure 2. Partial dependence plots of flight simulator performance versus predictors selected by Random Forest procedure (predictor value ranges as present in the sample). Black line/dots: Prediction averaged across 2,000 Random Forest fits. Gray lines/dots: Prediction of randomly selected individual fits to illustrate variability across fits. Note that the y-axes display only a small fraction of the outcome variable's theoretical range (0-14). ERI-Reward: Effort-Reward-Imbalance Reward Subscale. FFS: Full flight simulator. FTD: Flight training device. ALT: Alanine aminotransferase.

Auxiliary Appendix A: Methodological Details of Physiological Dysregulation Index Variable

Our physiological dysregulation index variable is based on all available measurements of the 18 biomarkers listed in Table A.I. The measurements were taken from the records of a total of 977 aeromedical examinations, conducted in the years from 2004 to 2016, that were available for 66 HEMS pilots. Of these, 51 with additional questionnaire and simulator rating data constitute the sample of the present study; records of 747 examinations are available for them. Since not all of the 18 markers had been measured in each examination, the total number of available measurements varies by marker (Table A.I). 16 of the 18 markers were also used as predictors on their own in the analysis (see main text). The remaining two were not used as the number of pilots with missing data was too large (≥33% of the sample, see Table A.I).

In the first step, we fitted for each biomarker an empty random intercept linear mixed model with the biomarker (log-transformed if necessary to better approximate normality) as the outcome and pilot as the grouping variable, i.e.

$$y_{it}^{(k)} = \beta^{(k)} + \gamma_i^{(k)} + \varepsilon_{it}^{(k)}$$

with $\varepsilon_{it}^{(k)} \sim N\left(0, {\sigma^{(k)}}^2\right), \gamma_i^{(k)} \sim N\left(0, {\tau^{(k)}}^2\right)$, and where k indexes the biomarker, i the pilot, and t the measurement occasion, and all within-pilot error terms ε and between-pilot error terms γ are assumed pairwise independent. For each combination of pilot i and biomarker k where at least one measurement of k was available, we stored the pilot's estimated average biomarker reading across all measurements, $\bar{y}_i^{(k)} = \hat{\beta}^{(k)} + \hat{\gamma}_i^{(k)}$, and for each k we stored the estimated pooled within-pilot standard deviation of k, that is, $\hat{\sigma}^{(k)}$ (all available measurements of k in the 66 pilots were used for this purpose).

Based on a review of literature and of clinical laboratory manuals, we determined upper and lower cutoffs c_u and c_l for each biomarker defining "healthy ranges" (Table A.I). Using the pilot averages $\bar{y}_i^{(k)}$ and pooled standard deviation $\hat{\sigma}^{(k)}$, we then calculated for each biomarker the estimated probability of a given pilot to have a biomarker reading outside the respective healthy range (e.g., systolic blood pressure >140 mmHg), assuming a normal distribution around the pilot average, i.e.,

$$P\left(y_{it'}^{(k)} \Big| \gamma_i^{(k)} > c_u \cup y_{it'}^{(k)} \Big| \gamma_i^{(k)} < c_l\right)$$

with $y_{it'}^{(k)} | \gamma_i^{(k)} \sim N(\beta^{(k)} + \gamma_i^{(k)}, {\sigma^{(k)}}^2)$, and where t' denotes a new measurement occasion. At this stage, a probability score was available for a given biomarker in all pilots where at least one measurement of that marker was available. We then applied the missForest imputation procedure as described in the main text to impute the missing probability scores (a total of 139=15.1% out of 18*51 data points), and finally summed the probabilities across all 18 biomarkers for each pilot to obtain the physiological dysregulation score, which can thus be interpreted as the pilot's expected number of biomarker readings outside the healthy range if one were to obtain measurements for all 18 markers in that pilot.

Table A.I. Details on individual biomarkers contributing to dysregulation score

Subsystem	Biomarker	Cutoffs	# of available measurements*	# of pilots with no available measurements*
Cardiovascular	Systolic blood pressure	> 140 mmHg	647	0
	Resting heart rate	> 80 bpm	665	0
	T axis deviation	> 60°	350	19
	QTc interval	> 450 ms	381	15
Metabolic	Body mass index	> 30 kg/m²	648	1
	Serum total cholesterol	> 6.21 mmol/l	498	3
	Serum HDL cholesterol	< 1.03 mmol/l	292	8
	Serum triglycerides	> 2.26 mmol/l	285	11
	Fasting plasma glucose	> 6.1 mmol/l	440	3
Liver	Alanine aminotransferase	> 60 U/l	200	14
	Aspartate aminotransferase	> 60 U/l	203	14
	γ-glutamyltransferase	> 60 U/l	148	22
Kidneys	Creatinine	> 115 µmol/l	230	8
Immune	White blood cell count	< 3.5 or > 10.5 * 10^3/µl	625	2
Blood	Hemoglobin	< 13.5 or > 18.0 g/dl	633	2
	RBC distribution width	> 15 %	537	4
Eyes	Intraocular pressure	> 21 mmHg	329	9
Ears	Hearing level at 3,000 Hz	> 25 dB HL	208	4

Note. QTc: Heart-rate corrected QT interval. HDL: High-density lipoprotein. RBC: Red blood cell. *In sample of pilots used in present study (N=51)

Auxiliary Appendix B: Analysis Variable Details

Table B.I. Description, sample items, and literature sources for individual analysis variables (binary yes/no variables in italics)

Variable category	Variable (inventory / scale name) [unit]	Brief description	Ref.
Outcome	Overall performance	Pilot's handling of malfunction scenarios "Transmission oil malfunction" and "Tail rotor drive failure" during simulator training flight (see main text)	-
Psychosocial work stressors	Emotional demands (COPSOQ II)	Requirement to deal with emotionally complex/difficult situations/interactions during work	25
	Social support (COPSOQ I)	Work-related support received by colleagues and superiors	18
	Work pace (COPSOQ II)	Tempo at which work tasks need to be conducted	25
	Work predictability	Degree to which day-to-day work tasks and events are perceived as expectable	34
	Role clarity (COPSOQ II)	Degree to which one's responsibilities at the job are known and understood	25
	Role conflict (COPSOQ II)	Inconsistency of different demands or responsibilities at the job	25
	Autonomy (TAA-KH-S)	Latitude at the job with regards to job contents, decision-making, and methods of carrying out one's work	9
	Supervisor Feedback (TAA-KH-S)	Explicit information received by supervisor about one's work behavior, performance, and results	9
	Procedural Justice	Degree to which one is given the chance to appropriately participate in organizational decision-making	17
	Effort at work (ERI)	Perceived demandingness of work / personal effort required at work	32
	Reward at work (ERI)	Perceived gratifying aspects of work, e.g. recognition, monetary rewards	32
Physical work stressors	Physical demands, general	Mechanical stresses imposed in general by work or working conditions, e.g. through heavy loads, repetitive movements or prolonged awkward stances	14
	Physical demands, headgear	Mechanical stresses imposed specifically by headgear worn during work	36
Psychosocial strain	Irritation (Irritation Scale)	Subclinical state of persistent irritability and rumination resulting from chronic discrepancies between one's (work) situation and personal goals	22
	Work engagement (UWES-9)	Degree of enthusiasm about / involvement in one's job. Defined as the opposite of burnout.	31
	Detachment from work (REQ)	Ability/willingness to mentally disengage from work during leisure time, to not occupy oneself with work-related issues at home	33
	Subjective well-being (WHO-5)	General sense of positive mood and vitality (not specifically with respect to work situations)	35
	Energy / Fatigue (WHOQOL-100)	Everyday level of energy (not specifically with respect to work situations)	40
Physical strain	# of body regions with pain	Number of body regions out of a total of 13 (e.g., neck, lower back) where subject reported "occasional" or "frequent" pain	19
Other work-related factors	Work hours per month	Average number of actual work hours per month	38
	Vacation days per year	Average number of actual days of vacation taken per year	8
	Day shift duty?	Working day shifts?	38
	Night shift duty?	Working night shifts?	38
	24 hour stand-by shift duty?	Working 24 hour stand-by shifts?	38
	Other shift type duty?	Working other types of shifts?	38
	Any limit on flight time?	Does any flight time limit per shift exist?	38
Medical risk markers	*Smoking?*	Smoking status, as reported by pilot at its most recent available assessment	1
	Any medication?	Medication use, as reported by pilot at its most recent available assessment	4
	Systolic blood pressure [mm Hg]	Systolic blood pressure, as measured at its most recent available assessment. Associated with cardiovascular morbidity and cognitive decline.	26
	Resting heart rate [beats per minute]	Resting heart rate, as measured at most recent available assessment. Associated with cardiorespiratory (un)fitness and cardiovascular morbidity.	15
	ECG QTc interval [msec]	Heart-rate corrected duration between onset of QRS complex to end of T wave in electrocardiogram, as measured at its most recent available assessment. Marker of myocardial repolarization abnormality.	28
	Body mass index [kg/m²]	Body mass index as measured at its most recent available assessment. Associated with cardiorespiratory (un)fitness, diabetes, and cardiovascular morbidity.	7
	Total cholesterol [mmol/l]	Serum total cholesterol concentration as measured at its most recent available assessment. Marker of lipid metabolism, associated with cardiovascular morbidity.	27
	HDL cholesterol [mmol/l]	Serum high-density lipoprotein cholesterol concentration as measured at its most recent available assessment. Marker of lipid metabolism. Low levels are associated with cardiovascular morbidity and diabetes.	27
	Triglycerides [mmol/l]	Serum triglyceride concentration as measured at its most recent available assessment. Marker of lipid metabolism, associated with diabetes.	11
	Fasting glucose [mmol/l]	Plasma fasting glucose concentration as measured at its most recent available assessment. Marker of glucose metabolism, associated with diabetes.	11

Table B.I. Description, sample items, and literature sources for individual analysis variables (binary yes/no variables in italics)

Variable category	Variable (inventory / scale name) [unit]	Brief description	Ref.
	Alanine aminotransferase [Enzyme units (U)/l]	Alanine aminotransferase activity as measured at its most recent available assessment. Marker of liver cell necrosis. May indicate liver conditions including alcoholic or nonalcoholic fatty liver disease.	3
	Aspartate aminotransferase [Enzyme units (U)/l]	Aspartate aminotransferase activity as measured at its most recent available assessment. Marker of liver cell necrosis. May indicate liver conditions including alcoholic or nonalcoholic fatty liver disease.	3
	Serum creatinine [μmol/l]	Serum creatinine concentration as measured at its most recent available assessment. Marker of kidney function (higher levels may be due to reduced renal clearance).	39
	White blood cell count [10^3/μl]	White blood cell count as measured at its most recent available assessment. Marker of immune system activity.	24
	Hemoglobin [g/dl]	Hemoglobin concentration as measured at its most recent available assessment. Low levels (anemia) lead to increased fatigability and reduced aerobic capacity. They may also be associated with reduced cognitive function.	29
	Red blood cell distribution width [%]	Coefficient of variation of volume of red blood cells, as measured at its most recent available assessment. Marker of perturbation in red blood cell maturation. Higher values often result from anemia, but may also be indicative of chronic inflammatory processes.	23
	Intraocular pressure [mm Hg]	Pressure inside the vitreous body of the eyes (averaged across both eyes), as measured at its most recent available assessment. High levels increase the risk for development of glaucoma.	20
	Hearing level at 3,000 Hz [dB Hearing Level]	Pure tone audiometry hearing level at 3,000 Hz (averaged across both ears).	37
	SCORE 6-month-risk [%]	Estimated probability to suffer a fatal cardiovascular event (e.g., myocardial infarction, stroke, sudden cardiac death) within the next 6 months. Estimation equation is derived from prospective cohorts of 11 European countries and is based on an individual's sex, age, country of origin, smoking status, systolic blood pressure, and total cholesterol level.	6
	Physiological dysregulation	Index of subclinical state of multi-organ functional decline. See main text and Auxiliary Appendix A.	2,5,30
Experience during simulator training session	Self-rated health	Subjectively experienced health status at the day of the simulator training session	21
	Task load during flight (NASA-TLX)	Subjectively experienced workload during simulator training flight (one item pertaining to one's own performance removed from the original scale)	13
	Simulator sickness	Degree of motion sickness induced during simulator flight	16
General pilot characteristics	Age [years]	Age at time of simulator training session	12
	# of real flight hours	Number of flight hours in real aircraft at time of simulator training session	12
	# of simulator flight hours	Number of flight hours in simulated aircraft at time of simulator training session	12
	Risk seeking (Self-control scale)	General tendency towards "adventuresome" behaviors and to enjoy risky activities	10
Other	*Full flight simulator?*	Simulator training flight in (motion capable) full flight simulator, vs. (motion incapable) flight training device. See main text.	-

Note. COPSOQ I/II: Copenhagen Psychosocial Questionnaire, Version 1/2. TAA-KH-S: Tätigkeits- und Arbeitsanalyseverfahren für das Krankenhaus, Selbstbeobachtungsversion [Activity and work analysis procedure for hospitals, self-report version]. ERI: Effort-Reward-Imbalance scale. UWES-9: Utrecht Work Engagement Scale, 9 item version. REQ: Recovery Experience Questionnaire. WHO-5: World Health Organization 5-item well-being index. WHOQOL-100: World Health Organization Quality of Life inventory, 100 item version. ECG: Electrocardiogram. HDL: High-density lipoprotein. SCORE: Systematic Coronary Risk Evaluation. NASA-TLX: National Aeronautics and Space Administration Task Load Index.

References to Auxiliary Appendix B

1. Ambrose JA, Barua RS. The pathophysiology of cigarette smoking and cardiovascular disease: an update. J Am Coll Cardiol 2004; 43(10):1731–1737.
2. Belsky DW, Caspi A, Houts R, Cohen HJ, Corcoran DL, Danese A, et al. Quantification of biological aging in young adults. Proc Natl Acad Sci U S A 2015; 112(30):E4104–E4110.
3. Burke MD. Liver function: test selection and interpretation of results. Clin Lab Med 2002; 22(2):377–390.
4. Chaturvedi AK, Craft KJ, Hickerson JS, Rogers PB, Canfield DV. Ethanol and drugs found in civil aviation accident pilot fatalities, 1989–2013. Aerosp Med Hum Perform 2016; 87(5):470–476.
5. Cohen AA, Milot E, Yong J, Seplaki CL, Fulop T, Bandeen-Roche K, et al. A novel statistical approach shows evidence for multi-system physiological dysregulation during aging. Mech Ageing Dev 2013; 134(3-4):110–117.
6. Conroy RM, Pyörälä K, Fitzgerald AP, Sans S, Menotti A, de Backer G, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. Eur Heart J 2003; 24(11):987–1003.
7. Eckel RH, York DA, Rossner S, Hubbard V, Caterson I, St Jeor ST, et al. Prevention Conference VII: Obesity, a worldwide epidemic related to heart disease and stroke: executive summary. Circulation 2004; 110(18):2968–2975.
8. Geurts SA, Sonnentag S. Recovery as an explanatory mechanism in the relation between acute stress reactions and chronic health impairment. Scand J Work Environ Health 2006; 32(6):482–492.
9. Glaser J, Seubert C, Hornung S, Herbig B. The impact of learning demands, work-related resources, and job stressors on creative performance and health. J Pers Psychol 2015; 14(1):37–48.
10. Grasmick HG, Tittle CR, Bursik RJ, Arneklev BJ. Testing the core empirical implications of Gottfredson and Hirschi’s General Theory of Crime. J Res Crime Delinq 1993; 30(1):5–29.
11. Grundy SM, Cleeman JI, Daniels SR, Donato KA, Eckel RH, Franklin BA, et al. Diagnosis and management of the metabolic syndrome: an American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement. Circulation 2005; 112(17):2735–2752.
12. Hardy DJ, Parasuraman R. Cognition and flight performance in older pilots. J Exp Psychol Appl 1997; 3(4):313–348.
13. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock PA, Meshkati N, editors. Human mental workload. New York, NY: Elsevier; 1988:139–183.
14. Hildebrandt VH, Bongers PM, van Dijk FJ, Kemper HC, Dul J. Dutch Musculoskeletal Questionnaire: description and basic qualities. Ergonomics 2001; 44(12):1038–1055.

15. Jensen MT, Suadicani P, Hein HO, Gyntelberg F. Elevated resting heart rate, physical fitness and all-cause mortality: a 16-year follow-up in the Copenhagen Male Study. Heart 2013; 99(12):882–887.
16. Johnson DM. Introduction to and review of simulator sickness research. Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences; 2005. Research Report 1832.
17. Kivimäki M, Elovainio M, Vahtera J, Ferrie JE. Organisational justice and health of employees: prospective cohort study. Occup Environ Med 2003; 60(1):27–33.
18. Kristensen TS, Hannerz H, Høgh A, Borg V. The Copenhagen Psychosocial Questionnaire – a tool for the assessment and improvement of the psychosocial work environment. Scand J Work Environ Health 2005; 31(6):438–449.
19. Kuorinka I, Jonsson B, Kilbom A, Vinterberg H, Biering-Sørensen F, Andersson G, et al. Standardised Nordic questionnaires for the analysis of musculoskeletal symptoms. Appl Ergon 1987; 18(3):233–237.
20. Mackenzie P, Cioffi G. How does lowering of intraocular pressure protect the optic nerve? Surv Ophthalmol 2008; 53 Suppl 1:S39–S43.
21. McDowell I. Measures of self-perceived well-being. J Psychosom Res 2010; 69(1):69–79.
22. Mohr G, Müller A, Rigotti T, Aycan Z, Tschan F. The assessment of psychological strain in work contexts. Eur J Psychol Assess 2006; 22(3):198–206.
23. Patel KV, Ferrucci L, Ershler WB, Longo DL, Guralnik JM. Red blood cell distribution width and the risk of death in middle-aged and older adults. Arch Intern Med 2009; 169(5):515–523.
24. Pawelec G, Goldeck D, Derhovanessian E. Inflammation, ageing and chronic disease. Curr Opin Immunol 2014; 29:23–28.
25. Pejtersen JH, Kristensen TS, Borg V, Bjorner JB. The second version of the Copenhagen Psychosocial Questionnaire. Scand J Public Health 2010; 38(3 Suppl):8–24.
26. Piepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, Catapano AL, et al. 2016 European Guidelines on cardiovascular disease prevention in clinical practice. Eur Heart J 2016; 37(29):2315–2381.
27. Reiner Z, Catapano AL, de Backer G, Graham I, Taskinen M-R, Wiklund O, et al. ESC/EAS Guidelines for the management of dyslipidaemias. Atherosclerosis 2011; 217(1):3–46.
28. Salles GF, Cardoso CR, Leocadio SM, Muxfeldt ES. Recent ventricular repolarization markers in resistant hypertension: are they different from the traditional QT interval? Am J Hypertens 2008; 21(1):47–53.
29. Schneider AL, Jonassaint C, Sharrett AR, Mosley TH, Astor BC, Selvin E, et al. Hemoglobin, anemia, and cognitive function: the Atherosclerosis Risk in Communities study. J Gerontol A Biol Sci Med Sci 2016; 71(6):772–779.

30. Seeman TE, McEwen BS, Rowe JW, Singer BH. Allostatic load as a marker of cumulative biological risk: MacArthur studies of successful aging. Proc Natl Acad Sci U S A 2001; 98(8):4770–4775.
31. Seppälä P, Mauno S, Feldt T, Hakanen J, Kinnunen U, Tolvanen A, et al. The construct validity of the Utrecht Work Engagement Scale. Multisample and longitudinal evidence. J Happiness Stud 2009; 10(4):459–481.
32. Siegrist J, Dragano N, Nyberg ST, Lunau T, Alfredsson L, Erbel R, et al. Validating abbreviated measures of effort-reward imbalance at work in European cohort studies: the IPD-Work consortium. Int Arch Occup Environ Health 2014; 87(3):249–256.
33. Sonnentag S, Fritz C. The Recovery Experience Questionnaire: development and validation of a measure for assessing recuperation and unwinding from work. J Occup Health Psychol 2007; 12(3):204–221.
34. Tetrick LE, LaRocco JM. Understanding, prediction, and control as moderators of the relationships between perceived stress, satisfaction, and psychological well-being. J Appl Psychol 1987; 72(4):538–543.
35. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: a systematic review of the literature. Psychother Psychosom 2015; 84(3):167–176.
36. van den Oord M. Prevention of flight-related neck pain in military aircrew [Dissertation]. Amsterdam: University of Amsterdam; 2012.
37. Wagstaff AS, Arva P. Hearing loss in civilian airline and helicopter pilots compared to air traffic control personnel. Aviat Space Environ Med 2009; 80(10):857–861.
38. Wagstaff AS, Sigstad Lie J-A. Shift and night work and long working hours – a systematic review of safety implications. Scand J Work Environ Health 2011; 37(3):173–185.
39. Wannamethee SG, Shaper AG, Perry IJ. Serum creatinine concentration and risk of cardiovascular disease: a possible marker for increased risk of stroke. Stroke 1997; 28(3):557–563.
40. World Health Organization. WHOQOL User Manual. WHO/MNH/MHP/98.4.Rev.1. Geneva: World Health Organization; 1998.

Auxiliary Appendix C: Additional Random Forest Variable Importance Output

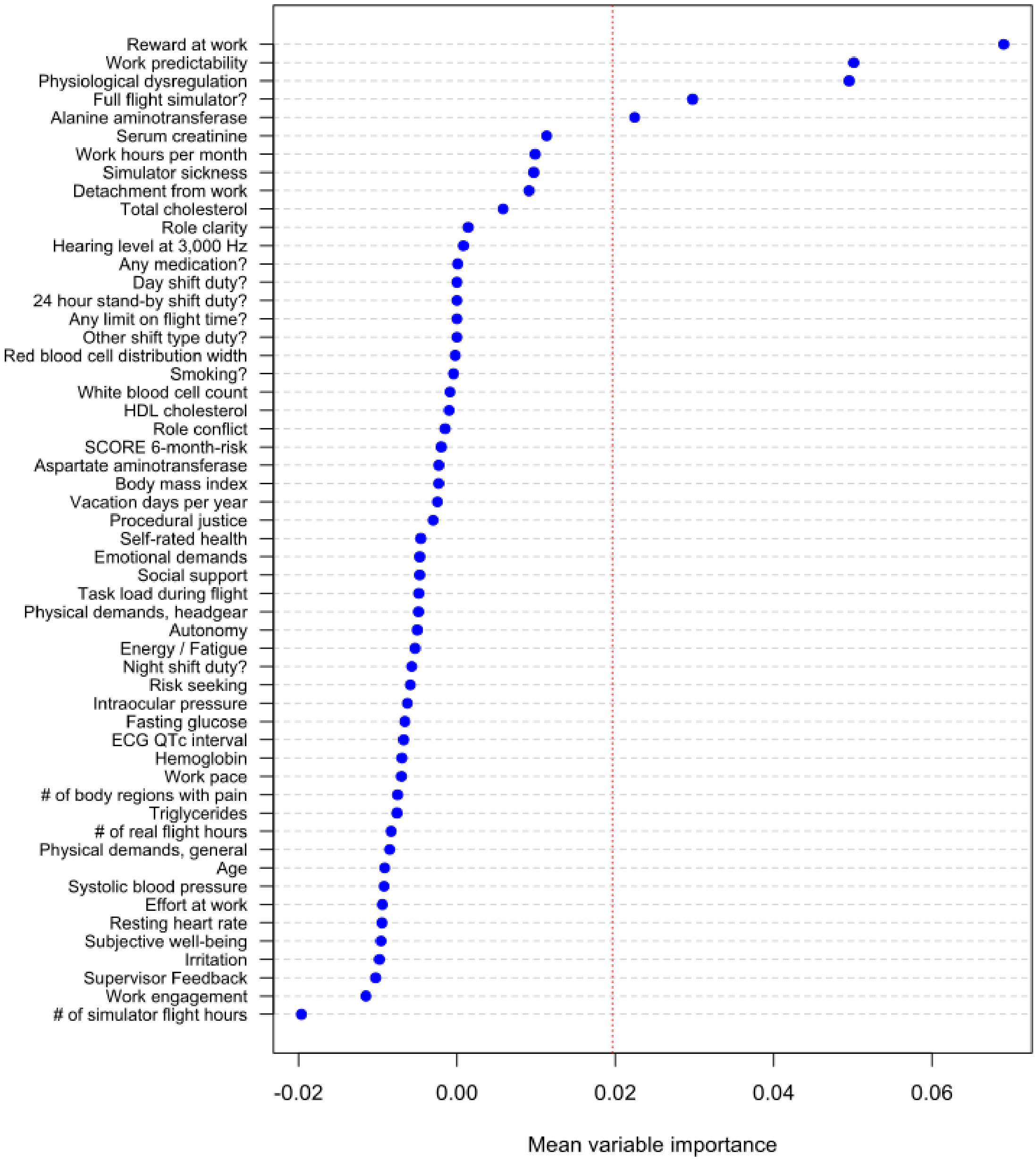


Figure C.1. Variable importance scores (averaged across 2,000 Random Forest fits) for all 54 predictor variables. Variables are binary if label ends with “?”, and quantitative otherwise. Red dotted line: Decision criterion for variable selection (absolute value of minimum mean variable importance).

Table C.I. Summary statistics for variable importance scores aggregated across 2,000 Random Forest fits

Variable	Untransformed variable importance							Rank-transformed variable importance (1 = highest, 54 = lowest)						
	Mean	SD	Min	1st Q	Median	3rd Q	Max	Mean	SD	Min	1st Q	Median	3rd Q	Max
Reward at work	0.069	0.016	0.020	0.058	0.069	0.080	0.120	1.4	0.7	6.0	2.0	1.0	1.0	1.0
Work predictability	0.050	0.015	0.003	0.039	0.050	0.060	0.116	2.6	1.2	16.0	3.0	2.0	2.0	1.0
Physiological dysregulation	0.050	0.015	0.003	0.039	0.048	0.060	0.113	2.6	1.2	14.0	3.0	3.0	2.0	1.0
Full flight simulator?	*0.030*	*0.011*	*-0.008*	*0.023*	*0.030*	*0.037*	*0.068*	*4.5*	*2.2*	*44.0*	*5.0*	*4.0*	*4.0*	*1.0*
Alanine aminotransferase	0.022	0.013	-0.025	0.014	0.022	0.031	0.070	6.4	5.2	54.0	7.0	5.0	4.0	1.0
Serum creatinine	0.011	0.008	-0.010	0.006	0.011	0.016	0.050	9.7	6.0	47.0	11.0	8.0	6.0	2.0
Work hours per month	0.010	0.008	-0.015	0.005	0.010	0.015	0.048	10.7	6.8	48.0	12.0	9.0	7.0	2.0
Simulator sickness	0.010	0.008	-0.016	0.004	0.009	0.015	0.043	10.5	6.7	53.0	12.0	9.0	7.0	3.0
Detachment from work	0.009	0.008	-0.017	0.004	0.009	0.014	0.040	11.6	8.2	50.0	12.0	9.0	7.0	3.0
Total cholesterol	0.006	0.009	-0.025	0.000	0.006	0.011	0.035	15.4	11.4	54.0	19.0	11.0	8.0	3.0
Role clarity	0.001	0.008	-0.026	-0.004	0.002	0.007	0.026	21.0	13.2	54.0	31.0	15.0	10.0	3.0
Hearing level at 3,000 Hz	0.001	0.009	-0.033	-0.005	0.001	0.007	0.030	22.8	14.6	54.0	34.0	17.0	10.0	4.0
Any medication?	*0.000*	*0.004*	*-0.014*	*-0.002*	*0.000*	*0.003*	*0.014*	*21.3*	*8.9*	*50.0*	*28.0*	*20.0*	*14.0*	*6.0*
24 hour stand-by shift duty?	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*20.1*	*2.5*	*29.5*	*21.5*	*20.0*	*18.5*	*13.0*
Day shift duty?	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*20.1*	*2.5*	*29.5*	*21.5*	*20.0*	*18.5*	*13.0*
Any limit on flight time?	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*20.1*	*2.5*	*29.5*	*21.5*	*20.0*	*18.5*	*13.0*
Other shift type duty?	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*20.1*	*2.5*	*29.5*	*21.5*	*20.0*	*18.5*	*13.0*
Red blood cell distribution width	0.000	0.007	-0.026	-0.005	0.000	0.004	0.021	23.4	12.9	54.0	33.0	20.0	12.0	4.0
Smoking?	*0.000*	*0.001*	*-0.005*	*-0.001*	*0.000*	*0.000*	*0.004*	*21.7*	*4.4*	*39.0*	*25.0*	*22.0*	*18.0*	*11.0*
White blood cell count	-0.001	0.010	-0.036	-0.007	-0.001	0.006	0.030	25.4	15.4	54.0	39.0	23.0	11.0	3.0
HDL cholesterol	-0.001	0.004	-0.020	-0.004	-0.001	0.002	0.014	23.8	9.9	53.0	31.0	24.0	15.0	6.0
Role conflict	-0.001	0.007	-0.025	-0.006	-0.001	0.003	0.023	25.5	13.1	54.0	36.0	25.0	13.0	4.0
SCORE 6-month-risk	-0.002	0.005	-0.021	-0.005	-0.002	0.001	0.015	26.1	10.6	54.0	34.0	26.0	17.0	5.0
Aspartate aminotransferase	-0.002	0.009	-0.033	-0.008	-0.002	0.003	0.027	27.4	14.8	54.0	40.0	27.0	13.0	2.0
Body mass index	-0.002	0.008	-0.033	-0.007	-0.002	0.003	0.031	27.2	14.3	54.0	39.0	26.5	13.0	4.0
Vacation days per year	-0.002	0.005	-0.020	-0.005	-0.002	0.001	0.015	27.2	10.5	53.0	35.0	27.0	17.0	5.0
Procedural justice	-0.003	0.005	-0.023	-0.006	-0.003	0.000	0.016	28.4	11.6	54.0	37.0	29.0	18.0	5.0
Self-rated health	-0.005	0.005	-0.028	-0.008	-0.004	-0.001	0.012	31.7	11.4	54.0	40.0	32.0	24.0	6.0
Emotional demands	-0.005	0.004	-0.027	-0.007	-0.004	-0.002	0.007	32.4	9.5	54.0	39.0	32.0	26.0	9.0
Social support	-0.005	0.007	-0.030	-0.009	-0.005	0.000	0.018	31.9	13.5	54.0	44.0	33.0	21.0	5.0
Task load during flight	-0.005	0.006	-0.029	-0.008	-0.004	-0.001	0.014	32.2	11.6	54.0	41.0	33.0	24.0	5.0
Physical demands, headgear	-0.005	0.007	-0.032	-0.009	-0.005	0.000	0.021	32.0	13.7	54.0	43.3	33.0	19.0	5.0
Autonomy	-0.005	0.005	-0.029	-0.008	-0.004	-0.002	0.009	32.7	10.6	54.0	41.0	33.0	26.0	8.0
Energy / Fatigue	-0.005	0.005	-0.027	-0.008	-0.005	-0.002	0.008	33.3	10.8	54.0	42.0	33.0	26.0	10.0
Night shift duty?	*-0.006*	*0.005*	*-0.024*	*-0.009*	*-0.005*	*-0.002*	*0.009*	*34.3*	*10.4*	*54.0*	*42.3*	*35.0*	*27.0*	*7.0*
Risk seeking	-0.006	0.006	-0.035	-0.009	-0.005	-0.002	0.013	34.3	11.4	54.0	43.0	35.0	27.0	6.0
Intraocular pressure	-0.006	0.005	-0.027	-0.009	-0.006	-0.003	0.004	35.5	9.7	54.0	43.0	36.0	29.0	9.0
Fasting glucose	-0.007	0.006	-0.029	-0.011	-0.006	-0.002	0.014	35.5	12.0	54.0	46.0	37.0	27.0	7.0
ECG QTc interval	-0.007	0.006	-0.030	-0.011	-0.006	-0.002	0.014	35.9	11.7	54.0	45.0	37.0	28.0	6.0
Hemoglobin	-0.007	0.006	-0.032	-0.011	-0.007	-0.003	0.013	36.1	11.8	54.0	46.0	38.0	28.0	7.0

Table C.I. Summary statistics for variable importance scores aggregated across 2,000 Random Forest fits

Variable	Untransformed variable importance							Rank-transformed variable importance (1 = highest, 54 = lowest)						
	Mean	SD	Min	1st Q	Median	3rd Q	Max	Mean	SD	Min	1st Q	Median	3rd Q	Max
Work pace	-0.007	0.006	-0.030	-0.011	-0.007	-0.003	0.008	36.4	11.3	54.0	46.0	38.0	29.0	7.0
# of body regions with pain	-0.008	0.006	-0.038	-0.011	-0.007	-0.004	0.007	37.7	10.4	54.0	46.0	39.0	31.0	9.0
Triglycerides	-0.008	0.006	-0.030	-0.011	-0.007	-0.003	0.010	37.5	11.1	54.0	47.0	39.0	30.0	9.0
# of real flight hours	-0.008	0.006	-0.037	-0.012	-0.008	-0.004	0.008	39.3	9.8	54.0	47.0	41.0	33.0	9.0
Physical demands, general	-0.008	0.007	-0.033	-0.013	-0.008	-0.004	0.009	38.9	11.3	54.0	48.0	41.0	32.0	8.0
Age	-0.009	0.006	-0.036	-0.013	-0.009	-0.005	0.010	40.2	10.4	54.0	49.0	42.0	33.0	8.0
Systolic blood pressure	-0.009	0.006	-0.037	-0.013	-0.009	-0.005	0.007	40.7	9.4	54.0	48.0	42.0	34.0	10.0
Effort at work	-0.009	0.006	-0.038	-0.013	-0.009	-0.005	0.006	40.9	9.6	54.0	49.0	43.0	35.0	9.0
Resting heart rate	-0.009	0.006	-0.036	-0.013	-0.009	-0.005	0.008	40.7	10.2	54.0	49.0	43.0	34.0	9.0
Subjective well-being	-0.010	0.006	-0.031	-0.013	-0.009	-0.006	0.006	41.4	9.4	54.0	49.0	43.5	36.0	9.0
Irritation	-0.010	0.006	-0.039	-0.014	-0.009	-0.005	0.008	41.3	9.9	54.0	49.3	43.0	35.0	9.0
Supervisor Feedback	-0.010	0.006	-0.036	-0.014	-0.010	-0.006	0.004	42.5	8.6	54.0	50.0	44.0	37.0	11.0
Work engagement	-0.011	0.007	-0.044	-0.016	-0.011	-0.007	0.005	43.6	9.3	54.0	51.0	46.0	38.0	10.0
# of simulator flight hours	-0.020	0.010	-0.056	-0.026	-0.019	-0.013	0.011	49.5	7.7	54.0	54.0	53.0	49.0	7.0

Note. Binary variables in italics. Q: Quartile.

4. Conclusions and outlook

When considering the results both of publication 1 submitted with this thesis, and of the as yet unpublished extended Age 60 Study results, one major conclusion is that the findings, based on different analytic approaches and data sources, consistently suggest the presence of a healthy worker survivor effect such that pilots in the oldest age group tend to show a more favorable, or at the very least comparably good, medical and psychological fitness profile as do middle-aged pilots. This calls into question the empirical basis of existing European legislation which bans pilots aged 60 and over from conducting single-pilot commercial air transport operations. In light of further developments in intelligent aircraft systems, single pilot operations in general are likely to become more prevalent and are an active field of research, including safety systems for automated detection and handling of pilot incapacitation (Vu, Lachter, Battiste, & Strybel, 2018).

Nevertheless, despite the study results being convergent and consistent with previous research (Aerospace Medical Association, 2004; Kay et al., 1994) as regards the arbitrariness of the age threshold of 60 years, further evidence will likely be demanded by European regulators in the course of the rulemaking process. One feasible avenue is to utilize the already established, very large airline pilot cohorts for whom cause-specific morbidity and mortality data are collected (Hammer et al., 2014; Linnersjö et al., 2011). A more ambitious approach, as suggested in publication 1, would be to establish a cross-country harmonized database of aeromedical findings, which could for example be used to re-calibrate (Blaha, 2016) existing cardiovascular risk scores to the pilot population or even to subpopulations such as HEMS pilots, or to construct new risk scores (not necessarily only regarding cardiovascular events as the outcome of interest, but also e.g. flight accidents or incidents) based on insights from the emerging field of prediction modelling (Moons, Kengne, Grobbee et al., 2012; Moons, Kengne, Woodward et al., 2012). Nonlinear modelling techniques, such as those applied in publication 1, play an important role in modern prediction models and would be an especially important consideration with regard to the effect of pilot age, since the empirical justifiability of decision thresholds like “age 60” strongly depends on the existence of corresponding threshold effects.

However, as also evidenced by the findings reported in publication 2, which identified two occupational stressors as significant predictors of simulator flight performance, and the analysis of stress and strain experienced by the HEMS pilots described in section 1.4.2, it is important to keep in mind the bigger picture, of which pilot age (let alone medical-cause sudden incapacitation) is only one aspect. Operational conditions are often of greater significance to flight safety than individual pilot characteristics. Machine learning methods are a promising approach to optimize this system layer, for example concerning prediction of weather characteristics (Williams, 2014) or automated recognition of high cognitive load situations in the cockpit (Harrivel et al., 2017). Going one step further, there are current efforts to establish intelligent systems as autonomous, co-pilot-like actors (“human-autonomy teaming”; Vu, Lachter, Battiste, & Strybel, 2018). Yet, even as visions of completely autonomous aircraft abound, it is quite likely (and probably for most, also desirable) that HEMS flights will continue to be operated by human pilots for a long time. The role of organizational factors and adequate working conditions, as emphasized by work design theory (Glaser, Seubert, Hornung, & Herbig, 2015; Humphrey et al., 2007) should therefore be given at least equal attention as technological solutions in future studies of flight safety in HEMS.

5. References

ADAC Luftrettung. (2017). Einsatzarten 2016 [Mission types in 2016]. Retrieved October 1, 2018, from https://stiftung.adac.de/_mmm/pdf/Einsatzarten%202016_294065.pdf

Aerospace Medical Association. (2004). The Age 60 Rule. *Aviation, Space, and Environmental Medicine*, *75*, 708–715.

Aerospace Medical Association. (2016). Pilot mental health: Expert working group recommendations – revised 2015. *Aerospace Medicine and Human Performance*, *87*, 505–507.

Aerossurance. (2018). Italian HEMS AW139 Inadvertent IMC Accident. Retrieved October 1, 2018, from http://aerossurance.com/helicopters/aw139-mountain-hems-iimc/

Agenzia Nazionale per la Sicurezza del Volo. (2018). *Relazione d'inchiesta. Incidente occorso all'elicottero AW139 marche di identificazione EC-KJT, in località Monte Cefalone, Lucoli (AQ), 24 gennaio 2017* [Investigation report. Accident of AW139 helicopter, registration code EC-KJT, at Monte Cefalone, Lucoli (AQ), 24 January 2017]. Retrieved October 1, 2018, from http://www.ansv.it/cgi-bin/ita/Relazione%20EC-KJT.pdf

Aherne, B. B., Zhang, C., & Newman, D. G. (2016). Pilot domain task experience in night fatal helicopter emergency medical service accidents. *Aerospace Medicine and Human Performance*, *87*, 550–556.

Arbeev, K. G., Ukraintseva, S. V., Bagley, O., Zhbannikov, I. Y., Cohen, A. A., Kulminski, A. M., & Yashin, A. I. (2018). "Physiological dysregulation" as promising measure of robustness and resilience in aging studies and new indicator of preclinical disease. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences.* Advance online publication. https://doi.org/10.1093/gerona/gly136

Arrighi, H. M., & Hertz-Picciotto, I. (1994). The evolving concept of the healthy worker survivor effect. *Epidemiology*, *5*, 189–196.

Baker, S. P., Grabowski, J. G., Dodd, R. S., Shanahan, D. F., Lamb, M. W., & Li, G. H. (2006). EMS helicopter crashes: what influences fatal outcome? *Annals of Emergency Medicine*, *47,* 351–356.

Barnett, K., Mercer, S. W., Norbury, M., Watt, G., Wyke, S., & Guthrie, B. (2012). Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet, 380,* 37–43.

Belsky, D. W., Caspi, A., Houts, R., Cohen, H. J., Corcoran, D. L., Danese, A., . . . Moffitt, T. E. (2015). Quantification of biological aging in young adults. *Proceedings of the National Academy of Sciences of the United States of America*, *112,* E4104–E4110.

Blaha, M. J. (2016). The critical importance of risk score calibration: time for transformative approach to risk score validation? *Journal of the American College of Cardiology*, *67,* 2131–2134.

Bledsoe, B. E. (2003). EMS myth #6. Air medical helicopters save lives and are cost-effective. *Emergency Medical Services*, *32*(8), 88–90.

Bledsoe, B. E., & Smith, M. G. (2004). Medical helicopter accidents in the United States: a 10-year review. *The Journal of Trauma*, *56,* 1325–1328.

Bledsoe, B. E., Wesley, A. K., Eckstein, M., Dunn, T. M., & O’Keefe, M. F. (2006). Helicopter scene transport of trauma patients with nonlife-threatening injuries: A meta-analysis. *The Journal of Trauma*, *60,* 1257–1265.

Blumen, I. (2009, February). *An analysis of HEMS accidents and accident rates.* Presentation at NTSB Public Hearing, Washington, D.C. Retrieved October 1, 2018, from https://www.ntsb.gov/news/events/Documents/NTSB-2009-8a-Blumen-revised-final-version.pdf

Blumen, I. J., & UCAN Safety Committee. (2002). *A safety review and risk assessment in air medical transport: Supplement to the Air Medical Physician Handbook*. Salt Lake City, UT: Air Medical Physician Association.

Branas, C. C., MacKenzie, E. J., Williams, J. C., Schwab, C. W., Teter, H. M., Flanigan, M. C., . . . ReVelle, C. S. (2005). Access to trauma centers in the United States. *JAMA, 293*, 2626–2633.

Brazier, J., Nicholl, J., & Snooks, H. (1996). The cost and effectiveness of the London helicopter emergency medical service. *Journal of Health Services Research & Policy, 1,* 232–237.

Brown, J. B., Rosengart, M. R., Billiar, T. R., Peitzman, A. B., & Sperry, J. L. (2017). Distance matters: Effect of geographic trauma system resource organization on fatal motor vehicle collisions. *The Journal of Trauma and Acute Care Surgery, 83,* 111–118.

Butler, B. (2014). Helicopter emergency medical services and weather-related accidents. *Air Medical Journal, 33,* 84–85.

Butler, D. P., Anwar, I., & Willett, K. (2010). Is it the H or the EMS in HEMS that has an impact on trauma patient mortality? A systematic review of the evidence. *Emergency Medicine Journal, 27,* 692–701.

Carchietti, E., Valent, F., Cecchi, A., & Rammer, R. (2011). Influence of stressors on HEMS crewmembers in flight. *Air Medical Journal, 30,* 270–275.

Chesters, A., Grieve, P. H., & Hodgetts, T. J. (2014). A 26-year comparative review of United Kingdom helicopter emergency medical services crashes and serious incidents. *The Journal of Trauma and Acute Care Surgery, 76,* 1055–1060.

Cohen, A. A., Li, Q., Milot, E., Leroux, M., Faucher, S., Morissette-Thomas, V., . . . Ferrucci, L. (2015). Statistical distance as a measure of physiological dysregulation is largely robust to variation in its biomarker composition. *PLoS ONE, 10,* e0122541.

Connell, L. J., & Reynard, W. D. (1995). Incident reports highlight hazards in EMS helicopter operations. *Helicopter Safety, 21*(4), 1–6.

Cowley, R. A., Mergner, W. J., Fisher, R. S., Jones, R. T., & Trump, B. F. (1979). The subcellular pathology of shock in trauma patients: Studies using the immediate autopsy. *The American Surgeon, 45,* 255–269.

Dambier, M., & Hinkelbein, J. (2006). Analysis of 2004 German general aviation aircraft accidents according to the HFACS model. *Air Medical Journal, 25,* 265–269.

de Wing, M. D., Curry, T., Stephenson, E., Palmieri, T., & Greenhalgh, D. G. (2000). Cost-effective use of helicopters for the transportation of patients with burn injuries. *The Journal of Burn Care & Rehabilitation*, *21*, 535–540.

Doucet, J., Bulger, E., Sanddal, N., Fallat, M., Bromberg, W., & Gestring, M. (2013). Appropriate use of helicopter emergency medical services for transport of trauma patients: guidelines from the Emergency Medical System Subcommittee, Committee on Trauma, American College of Surgeons. *The Journal of Trauma and Acute Care Surgery*, *75*, 734–741.

DuBois, G. (2005). The Age 60 Rule – it is time to defeat it. *Journal of Air Law and Commerce*, *70,* 319–352.

Era, P., Sainio, P., Koskinen, S., Ohlgren, J., Härkänen, T., & Aromaa, A. (2011). Psychomotor speed in a random sample of 7,979 subjects aged 30 years and over. *Aging Clinical and Experimental Research*, *23,* 135–144.

Evans, A. (2011, April). *Upper age limit for pilots*. Presentation at ICAO Regional Aviation Medicine Seminar for the NAM/CAR/SAM Regions, Mexico City. Retrieved October 1, 2018, from https://www.icao.int/NACC/Documents/Meetings/2011/AVMED2011/Day01-06-ICAO-Evans.pdf

Evans, A., Evans, S., & Harper, G. (2016). International regulation of medical standards. In D. Rainford & D. P. Gradwell (Eds.), *Ernsting's aviation and space medicine* (5th ed., pp. 357–372). Boca Raton, FL: CRC Press.

Federal Aviation Administration. (1959, December 5). Amendment 40-22 to §40.260 of 14 CFR Part 40: Maximum Age Limitations for Pilots. *Federal Register*, *24*, 9767–9768.

Fried, L. P., Tangen, C. M., Walston, J., Newman, A. B., Hirsch, C., Gottdiener, J., . . . McBurnie, M. A. (2001). Frailty in older adults: evidence for a phenotype. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, *56*, M146–M156.

Galvagno, S. M., Sikorski, R., Hirshon, J. M., Floccare, D., Stephens, C., Beecher, D., & Thomas, S. (2015). Helicopter emergency medical services for adults with major trauma. *The Cochrane Database of Systematic Reviews*, *12*, CD009228.

Glaser, J., Seubert, C., Hornung, S., & Herbig, B. (2015). The impact of learning demands, work-related resources, and job stressors on creative performance and health. *Journal of Personnel Psychology*, *14*, 37–48.

Habib, F. A., Shatz, D., Habib, A. I., Bukur, M., Puente, I., Catino, J., & Farrington, R. (2014). Probable cause in helicopter emergency medical services crashes: what role does ownership play? *The Journal of Trauma and Acute Care Surgery*, *77*, 989–993.

Hammer, G. P., Auvinen, A., de Stavola, B. L., Grajewski, B., Gundestrup, M., Haldorsen, T., . . . Blettner, M. (2014). Mortality from cancer and other causes in commercial airline crews: a joint analysis of cohorts from 10 countries. *Occupational and Environmental Medicine*, *71*, 313–322.

Hardy, D. J., & Parasuraman, R. (1997). Cognition and flight performance in older pilots. *Journal of Experimental Psychology: Applied*, *3*, 313–348.

Hardy, R., Lawlor, D. A., & Kuh, D. (2015). A life course approach to cardiovascular aging. *Future Cardiology*, *11*, 101–113.

Harrivel, A. R., Stephens, C. L., Milletich, R. J., Heinich, C. M., Last, M. C., Napoli, N. J., . . . Pope, A. T. (2017). *Prediction of cognitive states during flight simulation using multimodal psychophysiological sensing* (Report No. NF1676L-24736). Retrieved October 1, 2018, from https://ntrs.nasa.gov/search.jsp?R=20170001220

Hauser, W. A., Annegers, J. F., & Kurland, L. T. (1993). Incidence of epilepsy and unprovoked seizures in Rochester, Minnesota: 1935-1984. *Epilepsia*, *34*, 453–468.

Hickman, B. J., & Mehrer, R. (2001). Stress and the effects of air transport on flight crews. *Air Medical Journal*, *20*(6), 6–9.

Hinkelbein, J., Schwalbe, M., Neuhaus, C., Wetsch, W. A., & Genzwürker, H. V. (2011). Incidents, accidents and fatalities in 40 years of German helicopter emergency medical system operations. *European Journal of Anaesthesiology*, *28*, 766–773.

Humphrey, S. E., Nahrgang, J. D., & Morgeson, F. P. (2007). Integrating motivational, social, and contextual work design features: a meta-analytic summary and theoretical extension of the work design literature. *The Journal of Applied Psychology*, *92*, 1332–1356.

Huster, K. M., Müller, A., Prohn, M. J., Nowak, D., & Herbig, B. (2014). Medical risks in older pilots: a systematic review on incapacitation and age. *International Archives of Occupational and Environmental Health, 87*, 567–578.

Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction–job performance relationship: a qualitative and quantitative review. *Psychological Bulletin, 127*, 376–407.

Kay, E. J., Hillman, D. J., Hyland, D. T., Voros, R. S., Harris, R. M., & Deimler, J. D. (1994). *Consolidated Database Experiments Final Report: Age 60 Study, Part III* (Report No. DOT/FAA/AM-94/22). Retrieved October 1, 2018, from http://www.dtic.mil/dtic/tr/fulltext/u2/a286247.pdf

Kessler, C. (2015). Helicopter emergency medical service: motivation for focused research. *CEAS Aeronautical Journal, 6*, 337–394.

Linnersjö, A., Brodin, L.-A., Andersson, C., Alfredsson, L., & Hammar, N. (2011). Low mortality and myocardial infarction incidence among flying personnel during working career and beyond. *Scandinavian Journal of Work, Environment & Health, 37*, 219–226.

Lowsky, D. J., Olshansky, S. J., Bhattacharya, J., & Goldman, D. P. (2014). Heterogeneity in healthy aging. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, *69*, 640–649.

Martinussen, M. (2017). Pilot selection. An overview of aptitude and ability assessment. In R. Bor, C. Eriksen, M. Oakes, & P. Scragg (Eds.), *Pilot mental health assessment and support: A practitioner's guide* (pp. 23–39). New York, NY: Routledge.

Martinussen, M., & Hunter, D. R. (2010). *Aviation psychology and human factors*. Boca Raton, FL: CRC Press.

Mikkola, T. S., Gissler, M., Merikukka, M., Tuomikoski, P., & Ylikorkala, O. (2013). Sex differences in age-related cardiovascular mortality. *PLoS ONE, 8*, e63347.

Milot, E., Morissette-Thomas, V., Li, Q., Fried, L. P., Ferrucci, L., & Cohen, A. A. (2014). Trajectories of physiological dysregulation predicts mortality and health outcomes in a consistent manner across three populations. *Mechanisms of Ageing and Development, 141-142*, 56–63.

Moons, K. G., Kengne, A. P., Grobbee, D. E., Royston, P., Vergouwe, Y., Altman, D. G., & Woodward, M. (2012). Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*, *98*, 691–698.

Moons, K. G., Kengne, A. P., Woodward, M., Royston, P., Vergouwe, Y., Altman, D. G., & Grobbee, D. E. (2012). Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart, 98*, 683–690.

Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J., Cushman, M., . . . Turner, M. B. (2015). Heart disease and stroke statistics - 2015 update: a report from the American Heart Association. *Circulation*, *131*, e29–e322.

Müller, A., Prohn, M. J., Huster, K. M., Nowak, D., Angerer, P., & Herbig, B. (2014). Pilots' age and incidents in helicopter emergency medical services: a 5-year observational study. *Aviation, Space, and Environmental Medicine*, *85*, 522–528.

Nahrgang, J. D., Morgeson, F. P., & Hofmann, D. A. (2011). Safety at work: a meta-analytic investigation of the link between job demands, job resources, burnout, engagement, and safety outcomes. *Journal of Applied Psychology*, *96*, 71–94.

Newgard, C. D., Schmicker, R. H., Hedges, J. R., Trickett, J. P., Davis, D. P., Bulger, E. M., . . . Nichol, G. (2010). Emergency medical services intervals and survival in trauma: Assessment of the "golden hour" in a North American prospective cohort. *Annals of Emergency Medicine*, *55*, 235–246.e4.

Poguntke, P. (2012). Interview with F. Bruder, new CEO of ADAC Air Rescue: "Highest medical quality for people in distress". *AirRescue Magazine*, *2*(4), 14–15.

Radstaak, M., Geurts, S. A., Beckers, D. G., Brosschot, J. F., & Kompier, M. A. (2014). Work stressors, perseverative cognition and objective sleep quality: a longitudinal study among Dutch helicopter emergency medical service (HEMS) pilots. *Journal of Occupational Health*, *56*, 469–477.

Rasmussen, K., Røislien, J., & Sollid, S. J. (2018). Does medical staffing influence perceived safety? An international survey on medical crew models in helicopter emergency medical services. *Air Medical Journal*, *37*, 29–36.

Reason, J. (2000). Human error: models and management. *BMJ*, *320*, 768–770.

Ringburg, A. N., Buljac, M., Stolk, E. A., van Lieshout, E. M., van Beeck, E. F., Patka, P., & Schipper, I. B. (2009). Willingness to pay for lives saved by helicopter emergency medical services. *Prehospital Emergency Care, 13*, 37–43.

Rockwood, K., & Mitnitski, A. (2007). Frailty in relation to the accumulation of deficits. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, *62*, 722–727.

Rodenberg, H., Blumen, I. J., & Thomas, S. H. (2014). Air medical transport. In J. A. Marx, R. S. Hockberger, & R. M. Walls (Eds.), *Rosen's emergency medicine: Concepts and clinical practice* (8th ed., pp. 2442–2448.e2). Philadelphia, PA: Elsevier Saunders.

Rowe, J. W., & Kahn, R. L. (1997). Successful aging. *The Gerontologist, 37*, 433–440.

Salthouse, T. (2012). Consequences of age-related cognitive declines. *Annual Review of Psychology, 63*, 201–226.

Salthouse, T. A. (2004). What and when of cognitive aging. *Current Directions in Psychological Science*, *13*, 140–144.

Seeman, T. E., McEwen, B. S., Rowe, J. W., & Singer, B. H. (2001). Allostatic load as a marker of cumulative biological risk: MacArthur studies of successful aging. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 4770–4775.

Seplaki, C. L., Goldman, N., Glei, D., & Weinstein, M. (2005). A comparative analysis of measurement approaches for physiological dysregulation in an older population. *Experimental Gerontology, 40*, 438–449.

Stansfeld, S., & Candy, B. (2006). Psychosocial work environment and mental health–a meta-analytic review. *Scandinavian Journal of Work, Environment & Health*, *32*, 443–462.

Stuck, A. E., van Gorp, W. G., Josephson, K. R., Morgenstern, H., & Beck, J. C. (1992). Multidimensional risk assessment versus age as criterion for retirement of airline pilots. *Journal of the American Geriatrics Society, 40*, 526–532.

Swenor, B. K., Ramulu, P. Y., Willis, J. R., Friedman, D., & Lin, F. R. (2013). The prevalence of concurrent hearing and vision impairment in the United States. *JAMA Internal Medicine, 173*, 312–313.

Taylor, C. B., Stevenson, M., Jan, S., Middleton, P. M., Fitzharris, M., & Myburgh, J. A. (2010). A systematic review of the costs and benefits of helicopter emergency medical services. *Injury*, *41*, 10–20.

Unglücke der Luftrettung (2010, August 3). *Münchner Wochenanzeiger*. Retrieved October 1, 2018, from https://www.wochenanzeiger.de/article/100990.html

United Nations Department of Economic and Social Affairs Population Division. (2015). *World Population Ageing 2015* (ST/ESA/SER.A/390). Retrieved October 1, 2018, from http://www.un.org/en/development/desa/population/theme/ageing/WPA2015.shtml

US Department of Transportation. (1972, December). *Helicopters in emergency medical service: NHSTA experience to date* (DOT HS 820231). Retrieved October 1, 2018, from https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=17262

Vaidya, D., Becker, D. M., Bittner, V., Mathias, R. A., & Ouyang, P. (2011). Ageing, menopause, and ischaemic heart disease mortality in England, Wales, and the United States: modelling study of national mortality data. *BMJ*, *343*, d5170.

von Thaden, T., Wiegmann, D., & Shappell, S. (2006). Organizational factors in commercial aviation accidents. *International Journal of Aviation Psychology*, *16*, 239–261.

Vu, K.-P., Lachter, J., Battiste, V., & Strybel, T. Z. (2018). Single pilot operations in domestic commercial aviation. *Human Factors*, *60*, 755–762.

Wilkening, R. (2002). The Age 60 Rule: Age discrimination in commercial aviation. *Aviation, Space, and Environmental Medicine*, *73*, 194–202.

Williams, J. K. (2014). Using random forests to diagnose aviation turbulence. *Machine Learning*, *95*, 51–70.