Deriving risk based Target Levels of Safety (TLOS) for Autonomous Vehicle Systems

Vamsi K. Madasu and Kevin J. Anderson
SYSTRA Scott Lister
Level 39, 600 Bourke Street, Melbourne, Victoria
vmadasu@systra.com ; kanderson@systra.com

Abstract

A pragmatic and defensible set of risk-based Target Levels of Safety (TLOS) are defined for determining whether the level of risk is reasonably acceptable when evaluating the safety of mission critical systems. Although TLOS have been developed over the past 50 years by numerous stakeholders in different critical sectors (Hazardous chemicals, Rail, Aviation, Defence, Nuclear, Space, etc.), there is no consensus yet on risk-based TLOS for Autonomous Vehicle Systems.

This paper proposes TLOS for both ground (driverless trains and cars) and air (Unmanned Aircraft) based autonomous vehicle systems. The paper begins with a chronological review of the development of the various risk metrics and frameworks (such as ALARP, SFAIRP) across various industry sectors. A list of TLOS is then drawn up and studied to derive a risk baseline for the Australian public. Based on the unique nature of the Autonomous Systems (especially, those operating commercially) and socially accepted risk, a set of risk-based TLOS are defined for both Individual and Collective risk. Finally, these risk metrics are validated via prevalent aviation and ground transport safety regulations.

Keywords: Individual Risk, Collective Risk, Target Levels of Safety, Autonomous Vehicle Systems

1 Introduction

The approach to the reduction of risk posed by autonomous vehicle systems to the operators and the public should be a practical one. As with most transportation systems, risks associated with the operation of autonomous vehicles cannot be eliminated in totality. However, by establishing levels of risk which society considers acceptable and permitting the operation of only those systems which achieve these levels, regulatory authorities can effectively manage the risk associated with operation of such systems.

1.1 Outline of the paper

The introduction is immediately followed by Section 2 which reviews the different risk frameworks in practise, around the world. This is followed by a literature review of safety principles in Section 3. The concept of risk-based TLOS and the associated assumptions and derivations are given in Section 4. The risk-based TLOS are then presented for the three different types of autonomous vehicle systems, namely: Automated Trains; Driverless cars and Unmanned Aircraft Systems (UAS) in Section 5. These include references to all claims of fact, rationale behind assumptions and derivations of risk-based TLOS for each of the subject autonomous systems. Section 5 summarises the conclusions and provides some recommendations to the decision makers. Finally, the references are listed in Section 6 of the paper.

2 Risk Frameworks

The purpose of this section is to set the context for TLOS framework against the background of widely accepted safety principles. Safety is ‘freedom from unacceptable risk of harm’ [AS 61508, 2011] which presages a risk-based approach, often defined as necessary risk reduction .. ‘As Low As Reasonably Practicable (ALARP) [AS 61508, 2011] and its successor – So Far As Is Reasonably Practicable (SFAIRP) [ONRSR 2012].

The key question in determining whether a risk is ALARP is the definition of reasonably practicable. This term has been enshrined in the UK case law since the case of Edwards v. National Coal Board in 1949. The ruling was that the risk must be significant in relation to the sacrifice (in terms of money, time or trouble) required to avert it: risks must be averted unless there is a gross disproportion between the costs and benefits of doing so.’

In the absence of ALARP, attempts to reduce risks to zero could require infinite resources of time, effort and money. Benefit cost analyses can be used to demonstrate ‘gross disproportion’ but qualitative techniques can also be employed.

Risk based approaches are endemic to safety assurance and due diligence. We would list modern safety principles as follows:

- ‘Not less safe’
- ‘Compliance with standards’
- ‘Good practice’
- ‘SFAIRP’
- ‘Continuous improvement’

These concepts may be utilised in parallel or subsume one another.

2.1 ‘Not less safe’

The concept of ‘not less safe’ typically refers to incremental changes to an existing safety accreditation or certified safety-related system. For example, recent introduction of in-cab electronic authorities was tested against the question “*Has anything changed since the year 2000 approval which would invalidate that approval?*” [Advisian, 2015]. In the UK, at least, the term used is “*at least as safe*” for new equipment [INDG271, 2011].

2.2 ‘Good practice’

Claims of ‘best practice’ are rarely defendable. ‘Good Practice’ can be equated to ‘modern’ so long as any claim is attested through evidence.

2.3 ‘SFAIRP’

Safety legislation throughout Australia typically refers to ‘Safety-in-Design” and Occupational Health and Safety. For railways, the Rail Safety National Law and Regulations specify matters to be considered in Safety Management Systems (SMS) with a strong emphasis on risk reduction “So Far As Is Reasonably Practicable” (SFAIRP).

2.4 ‘Continuous improvement’

Acceptance of a “Not Less Safe” argument is often conditional on “Continuous Improvement”. In the railway Train Orders control system above, in-cab enforcement of Limit of Authority is under investigation for implementation when practicable. Taking account of lead times, testing and commissioning, we call this a risk ‘timeline’.

2.5 TLOS

The need for quantitative criteria to define socially acceptable Target Levels of Safety (TLOS) is critical to the seamless operation of all current and future autonomous vehicles in Australia and overseas.

A TLOS criterion is made up of both design and operational elements. Increasing levels of autonomy are expected to reduce the dependence on human interaction. So, it is appropriate that the operational/design balance to achieving the overall TLOS is reassessed and adjusted, where necessary, in favour of higher design standard.

For this paper, we drew on a literature review of current regulatory and risk frameworks in operation around the world (Section 2). Our review focused on design aspects (rather than operational and licensing issues) and more specifically, the use of quantified fatality risk targets. The purpose of this literature review was to provide the basis for a defensible position on setting safety targets for autonomous vehicle systems. To achieve this objective, we reviewed a range of documents which enumerate socially acceptable risk profiles for different types of industries, viz., Energy, Power, Aerospace, Defence and Rail, as well as the Safety Critical and Hazardous Chemicals sectors.

3 Literature Review

Several safety standards and handbooks were reviewed as part of the literature review on risk-based approaches to assuring the safety of critical systems. Apart from the AS 61508 standard, derived standards for various application sectors reviewed include AS 61511, EN 50126 and ISO 26262 [Ward, 2011].

Other relevant safety assurance standards reviewed include MIL-STD 882, DoD 00-56 and ARP 4761. Seminal references which detail the evolution of risk criteria were also studied [Lees 2011, HSE 1988, WA EPA, 1988 and HIPAP-4, 1989].

Based on the review of the above documents, the following salient points of interest can be noted on the history behind derivation of safety principles:

- Since World War II, the science of reliability and risk has flourished on the back of major unwanted world events
- Reliability theory post-war stemmed from a realisation as to just how many weapons programs failed.
- Space and aerospace achievements in the 1960’s were driven by systems engineering.
- In 1979, the Three Mile Island accident had lasting implications for consideration of human factors.
- The chemical industry became a focus for loss prevention following disasters such as Seveso, Bhopal, Flixborough, Piper Alpha
- A spate of major railway accidents in the 1990’s including Clapham Junction and Ladbroke Grove collisions and Kings Cross station fire heralded the introduction of risk-based Safety Management Systems (SMS).

In part response to some of the events above, the UK Health and Safety Executive [HSE, 1988] published the ‘Risk Triangle’. The same figure appears in AS 61508) and is reproduced here as Figure 1 with annotations by the authors, expressing views developed over many years as to the implied quantification of this triangle. This annotated Figure was first published in 1994 by VRJ Risk Engineers and later Risk & Reliability Engineers (1996 - 2005) with quantification based on tables in HSE and further work by WA EPA, NSW DoP and Vic VWA, tested through many assignments and training courses.

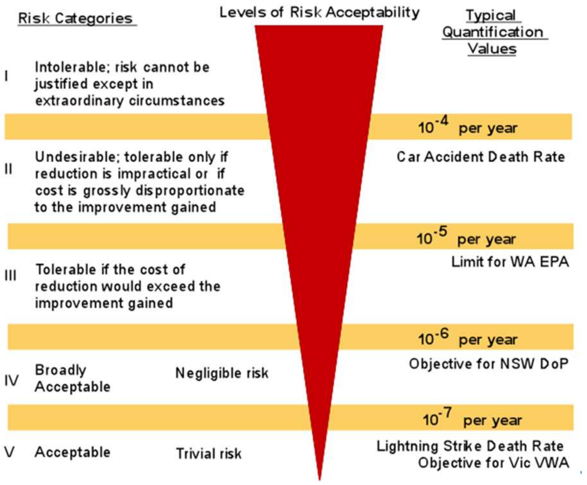
Risk Categories
Levels of Risk Acceptability
Typical Quantification Values
I
Intolerable; risk cannot be justified except in extraordinary circumstances
10^{-4} per year
II
Undesirable; tolerable only if reduction is impractical or if cost is grossly disproportionate to the improvement gained
Car Accident Death Rate
10^{-5} per year
III
Tolerable if the cost of reduction would exceed the improvement gained
Limit for WA EPA
10^{-6} per year
IV
Broadly Acceptable
Negligible risk
Objective for NSW DoP
10^{-7} per year
V
Acceptable
Trivial risk
Lightning Strike Death Rate
Objective for Vic VWA

Figure 1: Risk Triangle

Fig. 1 summarises the following categories of risk levels:

- Level I – Intolerable but for
- Level II – ALARP -Undesirable but for Gross Disproportion
- Level III – ALARP – Tolerable subject to benefit-cost comparison
- Level IV – Broadly acceptable
- Level V – Trivial

Note that UK's Health & Safety Executive [HSE, 1988] refers to the boundary between II and III as the benchmark representing the standard required to be met by new plant. In Australia, the states of NSW and WA issued Land Use Planning Guidelines [DEP, 2011] based on the same concept. Data on chances of fatality were central to arguments at that time as to tolerability of individual risk.

Table 1 combines NSW and HSE data regarding risks to individuals.

Table 1: Risk to individuals in NSW, 1989 and HSE, 1988

HIPAP, 1990	**NSW, per year**	**UK, per year**	**HSE, 1988**
		1.00E-2	Solo rock climbing
Smoking (voluntary)	5.00E-3		
Cancer (population)	1.80E-3		
		1.00E-3	'High risk' industry
Alcohol (voluntary)	3.80E-4		
Motor vehicle (traveller)	1.45E-4	1.00E-4	Motor vehicle (traveller)
Train (traveller)	3.00E-5		
Aeroplane (traveller)	1.00E-5		
		1.00E-5	'Safe' industry
Accidents at home	1.10E-4		
Falls	6.00E-5		
Pedestrian struck by vehicle	3.50E-5		
Professional drivers	3.50E-5		
Homicide	2.00E-5		
Fire /electrocution	1.30E-5		
		1.00E-6	Home gas fire /explosion
Lightning strike	1.00E-7	1.00E-7	Lightning
Meteorite strike	1.00E-9		

We have taken 'One in a million' to represent 'broadly acceptable' individual risk. In exponential notation, this is 1.00E-6 per year. See also HIPAP-4 [DEP, 2011] with regard to criteria for individual exposure per year to heat radiation, explosion overpressure and toxic exposure for:

• Hospitals, schools, child-care, old age housing - 0.5 in a million per year	5×10^{-7}
• Residential, hotels, motels, tourist resorts	1×10^{-6}
• Commercial developments including retail centres, offices and entertainment centres	5×10^{-6}
• Sporting complexes and active open space	1×10^{-5}
• Industrial	5×10^{-5}

HIPAP-4 also quotes HSE view that risk to a member of the public might be regarded as acceptable, as opposed to tolerable, at 1.00E-6 per year. 100 in a million (1.00E-4) is taken to represent 'Intolerable – except in extraordinary circumstances'. HIPAP note that HSE suggest limits of tolerable risk to a worker 1.00E-3 per year and to a member of the public 1.00E-4 per year

A particularly thing to note here is…'*to learn that over 5,000 people are killed each year by traffic does not prevent us from using the roads, though it warns us to be cautious'* [HSE, 1988]. The comparison in this paper of 'Intolerable' to the car accident death rate stemmed from the above examples quoted in HIPAP No. 4. Note also that governments tend to favour an overall road transport view (where there are X deaths per year), as compared to the individual vehicles/systems view favoured by manufacturers. Given an estimate of the exposed population, the numbers are relatively similar, as further discussed below.

Exponentially, in the middle between 'Tolerable' and 'Undesirable, lies 10 to 30 in a million (1.00E-5 to 3.00E-5) – our region of focus for this analysis. Similarly, Fig. 4 of HSE, 1988 drew a middle line here as the 'Benchmark representing the standard required to be met by new plant.

An underlying assumption in this analysis is that society and politicians accept risk-based methods. Furthermore, it is stated in HIPAP, supported by data on annual risk from all causes in the UK that if a risk from a potentially hazardous installation/operation is below most risks being experienced by population age groups, then such risk (1.00E-6 per year) may be tolerated. This is not the same as background risk.

It is also accepted in risk management literature that the quantitative risk criteria for public are normally more than that for workers (due to the voluntary nature of the work). Refer to Table 1 for a range from 1.00E-5 to 1.00E-3 for 'Safe' and 'High risk' industries respectively.

The maximum tolerable level of risk therefore depends on the occupation of a worker. Setting a level of 1.00E-4 per year implies that extraordinary circumstances exist for rock climbers, helicopter pilots and other 'high risk' industries. Another view based on the road toll is that "*if your job is more dangerous than your journey to work, your employer has a case to answer*". A further approach is to delineate the ALARP zones by one order of magnitude each. Refer Figure 1.

Collating the above views suggests to use a target TLOS figure of 1.00E-5 per year for individual risk to workers and 1.00E-6 to the public.

Societal or group risk is also a vexed problem. Society seems to be more averse to the risk of multiple fatalities in one event – fire, explosions and toxic releases, mid-air aircraft collisions, for example. Train collisions and derailment, road accidents and unmanned UAV represent lesser degrees of societal risk.

For these, we have coined the term 'collective' risk, as further discussed below, in preference to the traditional F-N plots of cumulative frequency F of N or more fatalities.

4 Risk-based TLOS

Risk is combination of the probability of occurrence of harm and the severity of that harm [AS 61508, 2011]. In technical terms, risk is a metric that accounts for both consequence and probability over a specified interval of exposure.

4.1 Risk Metrics

The definitions of some common risk metrics are as follows:

- Individual risk: Individual fatality risk is the risk of 'death to a person at a particular point' [DEP, 2011]. This is the risk experienced by a single individual in a specific time-period at a given location. It reflects severity of the hazard (consequence) and amount of time, the individual is in proximity to the risk (likelihood). Individual Risk describes the risk to a certain person of becoming a casualty, at a certain location. We have expressed this on a per hour basis
- Societal or Group risk: these reflect societal concerns as to the occurrence of multiple fatalities in a single event [HIPAP-4].
- Collective risk: In general terms, this is defined as the risk experienced by a group of people exposed to the hazard, often expressed as a relationship between frequency of an event (likelihood) and the number of people affected by that event (consequence). Collective risk is a single number that reflects the total aggregated risk posed by the entire activity. In other words, it combines all the individual risks, and accounts for multiple casualties from a single event at a given time and place. We have expressed collective risk on a per annum basis as individual risk per hour times the annual exposure to the activity.

In this paper, we have adopted the Individual and Collective Risk metrics to propose the TLOS for Autonomous Vehicle Systems.

4.1.1 Individual risk

Individual risk is particularly useful where certain individuals are exposed to a higher risk than others. For example, the individuals whose homes are at the end of a railway line or directly under a persistently loitering UAS; may be exposed to a risk that exceeds acceptable levels (and therefore may require risk treatment). If, however, an activity exposes all individuals to broadly the same level of risk, then 'individual risk' is a less useful metric.

4.1.2 Collective risk

'Collective risk', on the other hand, as defined, is always useful to regulatory authorities as it describes, via a single number, the total risk for an activity. When compared against a benchmark (such as a collective risk TLOS), it provides a measure of the acceptability of the overall risk of an activity

For the purposes of SFAIRP, cost-benefit analyses of necessary risk reduction assume a valuation of a fatality statistically averted. A figure of $1M was adopted in the 1990s which increased to $3M, a decade later. Assuming a figure of $10M today for comparative purposes, an individual risk TLOS of 1.00E-6 per annum carries negligible value in terms of willingness to spend. It is the aggregate or collective risk that carries weight. Further assuming a road toll individual risk safety target of 50 chances per million years and then applying a risk reduction of 90%, we obtain a residual risk of 5 in a million (5.00E-6 per year) which when averaged over a population of 25 million is 1,088 less fatalities per annum – or in cost terms, a saving of $10.1B. This sets the tone for the application of the SFAIRP criterion of 'gross disproportion' which in turn drives the TLOS assumptions.

4.2 TLOS approach based on risk

The TLOS approach for establishing a risk criteria framework for autonomous vehicle operations is based on setting an overall safety objective for the autonomous vehicle systems within the broad context of defined activities and operating environments. This process consisted of four disparate steps:

- STEP 1: A literature review was conducted of current regulatory and risk frameworks in operation around the world. This literature review focused on the use of quantified casualty/fatality risk targets. The purpose of this literature review was to provide the basis for a defensible position on setting TLOS for autonomous vehicle operations. To achieve this objective, we reviewed a range of documents which enumerate socially acceptable risk profiles for different types of industries, such as, Energy, Power and Defence, as well as the relevant Road, Rail and Aerospace sectors.
- STEP 2: Quantitative risk criteria for a fatality (serious injury /loss of life) are usually expressed as an annualised frequency of occurrence. Alternatively, risk can also be expressed as per exposed flight or driving hour probability of failure. We have defined individual risk in fatalities per exposed hour and collective risk in per annum terms. The following assumptions were made to convert all risk criteria metrics to align with these definitions to a standard scale:
  - An exposure time of 1760 hours per year (40 hours/week × 44 weeks/year) is assumed for workers to convert from individual risk per annum to individual risk per hour.
  - Similarly, an exposure time of 8760 hours per year (24 hours/day × 365 days/year) is assumed for members of the public as they a near universal exposure to different modes of transport.
- STEP 3: The normalised data obtained from literature review was then analysed for deriving the most common individual and collective risk criteria for workers and members of the public, across the surveyed industries. The purpose of selecting the most common criteria was to establish target levels of safety which have broad based understanding, legal precedence, and support within the technical community.
- STEP 4: The relevant TLOS criteria were obtained by converting the most common criteria for individual risk per annum to a per hour basis.

The above process is an extension of the first author's previous work in the UAS field [AMOG, 2012] and has been adopted here for deriving TLOS for other types of autonomous vehicle systems. The TLOS are mostly based on the country-specific data obtained for fixed hazardous installations such as chemical/nuclear power plants.

The proposed quantitative risk criteria are best put into practice when used as safety targets rather than absolute threshold levels. Given that we have followed HSE's view regarding the mid-point, exceedance within one order of magnitude could still be tolerable, albeit undesirable and subject to the test of gross disproportion in relation to the SFAIRP test.

5 TLOS for Autonomous Vehicle Systems

5.1 Autonomous Vehicles

An autonomous vehicle is characterised by the absence of a driver/pilot on-board the vehicle. Ground-based autonomous vehicles such as driverless cars and automatic trains may still carry passengers.

Some commentators suggest that the term 'autonomous' should only be referred to vehicles carrying occupants. Whereas, in the absence of any human occupants on-board the vehicle, the terms unmanned/uninhabited/driverless should be used. In principle, we have followed this terminology about the UAS definition. However, 'driverless' refers to inhabited vehicles.

It is noted that UAV may also carry passengers in future. However, this paper considers only unmanned UAV i.e. neither pilot nor passengers i.e uninhabited. Unmanned Aircraft Systems (UAS) where no person is on-board the air vehicle imply that risk of any collateral damage is limited to third parties and property on the ground including workers and members of the public.

5.1.1 Assumptions

The risk posed by autonomous vehicle systems depends not only on the reliability of the system but also on the characteristics of the operating area (i.e. population distribution along rail/airspace corridors). Any TLOS for such systems, therefore, must take both aspects into account if they are being designed to protect the public without unnecessarily inhibiting the development and integration of autonomous technologies in the ground and air transport realms.

Several critical assumptions were made while deriving the TLOS for Autonomous Vehicle Systems. Some of the key assumptions are:

- Members of the public should have a lesser exposure to risk than that of workers. This is because workers by their occupation are aware of (or should be aware of) the hazards and risks involved with such operations. Conversely, the public has no relation to the operation of autonomous vehicles and should therefore have a lower exposure to risk than workers.
- Autonomous vehicles should not introduce a level of risk greater than that already tolerated by society. This is based on our view that the risk to general population resulting from an operation or a facility should not exceed the sum of fatalities resulting from other accidents to which members of the public are generally exposed. We note that regulators and responsible authorities strive to improve safety such that the risk to life is either reduced or at least doesn't increase.
- Autonomous vehicles pose no more than 10% of the annual risk of fatality to workers and members of public. The risk posed by these systems should not exceed the sum of fatality risks (aggregate risk) resulting from other accidents to which members of the general population are already exposed to. Furthermore, the risk posed by UAS operations was reasonably considered to be no more than 10% of this aggregated risk.

5.2 Automated trains

Automated trains have been deployed in many countries for some years, now. Indeed, there are few new Metropolitan train lines that are not either partially or fully automated.

Grades of Automation are specified by the standard IEC 62290 [IEC 62290, 2014] ranging from fully automatic including door closing, obstacle detection and emergency situations through to having on-board staff or drivers handling these situations. The European Railway Agency (ERA) regularly reports on the achievement of National Reference Values (NRV) and Common Safety Targets (CST) for different rail systems in Europe. ERA's statistics are principally based on the concept of 'not less safe'.

In deriving TLOS for automated trains, we utilise the same approach as the European Railway Agency which focuses on National Reference Values (NRV) and Common Safety Targets (CST) which in turn are based on normalised occurrence reporting.

Table 2: Proposed TLOS for Automated Trains

	Workers	General Public
Individual risk	3.5E-08 fatalities per hour	3.5E-09 fatalities per hour
Collective risk	3E-05 fatalities per annum	3E-06 fatalities per annum

The main train driver safety hazards are Overspeed and Signal Passed at Danger (SPAD). These hazards are usually mitigated by engineering controls such as Automatic Train Protection (ATP) or other technologies, including the use of Global Positioning Systems (GPS) to improve driver situational awareness.

Driverless trains are further planned for Sydney's North West Rail Link (NWRL). Given modern Safety Management Systems (SMS) introduced in the wake of Glenbrook and Waterfall Rail Accident Inquiries, it is expected that the new trains will be more than 'not less safe' than existing rolling stock.

5.3 Driverless cars

The media routinely run speculative articles about the advent of driverless cars in the next decade. Requirements for Driverless Cars (DC) were examined by Anderson and Boughton [ASSC 2016] addressing components as depicted in Figure 2:

- Information inputs – GPS, Radar, Lidar, Odometry and Image processor are given in orange circles
- Processing – mission computer, knowledge base and artificial intelligence are blue-green clouds
- Outputs – Steering, accelerator and brakes are pink hexagons.

Anderson and Boughton (2016) considered risk reduction strategies for driverless cars using the 'So Far As Is Reasonably Practicable' (SFAIRP) principle. The following questions were asked at the ASSC2016 workshop conducted by the authors:

1 Will Rear end collisions still occur as DC behaviour shapes human behavior and vice versa?

2 Will Manufacturers /programmers will be held to a higher standard than drivers?

3 If the lesser evil has to be chosen i.e. will we let one person die to save a school bus?

4 Are Driverless cars obligated to save their passengers irrespective of tradeoff – loss potential?

5 Will there be mega-accidents as cars are networked, vulnerable to hacking?

6 We cannot continue to kill 30,000 people US alone (Aus 1,209 fatalities) ?

The workshop answered yes to each of these Questions. The point of the exercise was to deliberately test ethical questions against a professional audience. For example, a professional safety engineer would respond to Question 3 by saying that our objectives should be to engineer the systems so that we avoid getting to that choice (notwithstanding the argument about residual risk)

While Question 5 could be regarded as worthy of the tabloid press, were it not for the expectation that the industry would provide appropriate Safety Integrity Levels (SIL or RSIL). The automotive industry is aware of, and seeking to address, the security challenges of networked vehicles.

Reviewing these answers in the light of Target Levels of Safety - Questions 1 and 5 demonstrate that hazards cannot be eliminated and residual consequences remain. At Question 2, legislation is required to clarify the obligations of manufacturers given the difficult ethical trade-offs revealed at conclusion 3 and conclusion 4. Question 6 suggests a 'not less safe' comparison with the current road toll will not be tolerated.

Principles of 'continuous improvement', 'good practice' and 'compliance with standards' suggest a risk reduction target to equate TLOS to that ALARP midpoint.

In 1995 there were 2,017 road fatalities in Australia – an average for all individuals of 111.6 chances per million years (itself an improvement on the 1989 figure of 145 chances per million years). By 2015, this figure had fallen to 1,209 fatalities at an average of 50.8 chances per million years.

Accordingly, we recommend to set the Driverless Car TLOS as a fraction of the road toll – Given that the road toll itself has been reduced in the last 30 or so years, this would equate to fatality risks experienced by professional drivers – couriers, taxi drivers etc. of say, around 3.00E-5 per year.

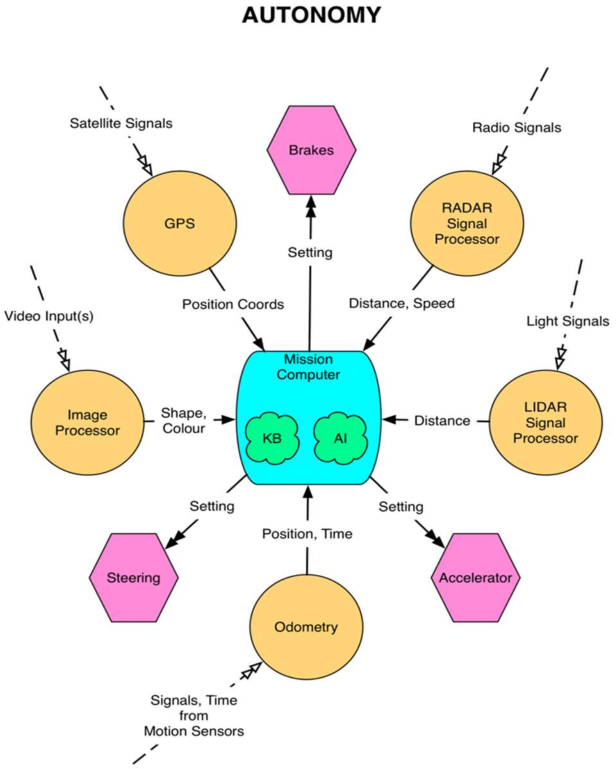

Figure 2: Driverless Car requirements

Over the next twenty years or so, significant reductions can be expected, better than what has already been achieved, continually reducing the 'not less safe' target for driverless cars. Monitoring of technology as to what is reasonably practicable should accelerate these projections.

'Towards Zero' is a vision of nationwide transport safety authorities for a future free of deaths and serious injuries on Australia roads. It is argued that this is necessary to ensure a safe road system in place. 'Zero harm' is a catchcry of many corporate businesses too. But we do not think that this is realistic. In fact, we believe that there could be an expectation of significant further risk reduction over the next 20 years towards a baseline acceptable risk. As above, we further believe that monitoring of technology as to what is reasonably practicable should (if not must) accelerate these projections.

For driverless cars, we propose a TLOS of 10% of the Australian road toll, i.e. 5 chances per million years' individual risk to the travelling public. This figure is then extrapolated to obtain other risk metrics (see Tables below). Possibly, there many other detailed reports on this topic for the automotive industry, but results are not readily available. In fact, the derivative of AS 61508 standard – ISO 26262 has only recently been updated and approved [Ward, 2011].

Table 3: Proposed TLOS for Driverless Cars

	Workers	General Public
Individual risk	6E-09 fatalities per hour	6E-10 fatalities per hour
Collective risk	5E-05 fatalities per annum	5E-06 fatalities per annum

Over the next decade, accelerating autonomous driving technology, including advances in artificial intelligence, sensors, cameras, radar and data analytics, are set to transform not only how we drive (or, indeed, are driven), but the notion of car ownership itself.

5.4 TLOS for Unmanned aircraft systems

An Unmanned Aircraft System (UAS) is an aircraft and its associated elements which are operated with no pilot on board [ICAO RPAS Manual, 2015].

Risk assessments for UAS operations have the same goal (public safety) as risk assessments for manned aircraft but must consider the unique flexibility afforded by unmanned aircraft. The risk associated with operating aircraft may be divided between three primary groups: the crew and passengers aboard the primary aircraft, the crew and passengers of other nearby aircraft and people and property on the ground.

When considering the safety of manned aircraft, if the first group on the primary aircraft is always assumed to be safe. If so, the other two will largely be safe as well (an exception being some residents under the immediate flight path). Manned aircraft must be extremely reliable because any crash is a threat to all the people on-board the aircraft. The area in which the aircraft is operating does not affect the need for reliability. However, when the crew and passengers are not present in the aircraft, the established approach of focusing on the safety of the people on-board the aircraft is no longer applicable. Hence the risk posed by UAS operations is either to people/property on the ground or to other airspace users.

Quantitative means of assessing risk posed by various UAS [AMOG, 2012] is important to the development of effective risk mitigation strategies and the formulation of Concept of Operations. Qualitative assessments are also appropriate in some contexts [JARUS, 2015]. However, the scope of this paper is limited to the evaluation of risk posed by a UAS to ground populations.

There are several issues involved while evaluating the risk posed by UAS to ground populations: UAS may impinge on controlled airspace, either accidently, deliberately or through a form of exceptional clearance; military or even civilian aircraft may exit controlled airspace into uncontrolled airspace; all leading to mid-air collision scenarios, outside of segregated airspace and near large ground populations. As such, it is very difficult to predict the number of people who would be exposed to UAS ground crashes. For this analysis, we have assumed UAS to be operated in segregated airspaces, away from large population centres.

UAS risk analysis typically computes the maximum individual risk as the highest probability any given individual has of suffering a serious injury or worse (i.e. becoming a casualty) because of a UAS crash. The consequence implicit in any individual risk is an adverse outcome for a single individual, thus individual risk is a quantity that is bounded by zero and one. In other words, the maximum individual risk from an event is always bounded between no possibility and absolute certainty of an adverse consequence.

In contrast, collective risk is defined here as the risk of an adverse outcome among a group of individuals. This is distinct from the societal risk concern of multiple fatalities. Collective risk can be expressed in terms of expected values: the consequences that occur because of an event if the event were to be repeated many times in the same locality. Collective risk is therefore analogous to an estimate of the average number of people injured by a crash (i.e. site occupancy), while individual risk would be the likelihood of an individual at a location being injured by the crash.

The proposed set of quantitative risk criteria for UAS are illustrated in Table 4.

Table 4: Proposed TLOS for UAS

	Workers	General Public
Individual risk	6E-08 fatalities per hour	6E-09 fatalities per hour
Collective risk	5E-04 fatalities per annum	5E-05 fatalities per annum

6 Conclusions and Recommendations

This paper proposes quantitative risk-based (Individual and Collective Risk) TLOS for autonomous vehicle systems.

The key to ensuring the safe employment of autonomous vehicles is not by just adopting subjective thresholds to manage the public's perception of risk, or developing safety tools for enforcement or even risk assessments. Rather, the solution lies in standardising the process by which risks are assessed and undertaking efforts to reduce the gap between real versus perceived risks. The proposed quantitative risk criteria are therefore best put into practice when used as safety targets rather than absolute threshold levels. An exceedance within one order of magnitude could still be tolerable subject to the SFAIRP test.

In this paper, we have steered away from setting a catastrophic or societal risk limit which is generally presented as curves on F-N plots, where F is the cumulative frequency of N or more fatalities and N is the number of fatalities.

Although all accidents resulting in a fatality are a cause for concern, society normally tends to be more alarmed when multiple fatalities occur in a single event. Whilst such low-incidence high-effect events might represent a very small risk to an individual, they may be viewed as unacceptable when many people are exposed to accident. However, the probability of a large population being exposed to such an event is negligible given the fact that currently autonomous systems are well segregated from conventional vehicles. Hence, the likelihood of a multiple fatality accident involving an autonomous vehicle is extremely low and therefore, we believe that the need for a catastrophic risk limit is not warranted at this stage.

7 References

Advisian (2015). Functional safety assessment of Train Management and Control System (TMACS). *Technical Report*.

AMOG Consulting (2012). A UAS Quantitative Risk Criteria – Development of Safety Targets for ADF Category 2 UAS. *Technical Report*.

Anderson, K. and Boughton, C. (2016). Facilitated discussion on safety and security issues for Driverless Cars. *Australian System Safety Conference 2016*

Anderson, K. et al. (1996). Risk & Reliability – An introductory text.

AS 61508.4-2011 (2011). Functional safety of electrical/electronic/ programmable electronic safety-related systems. *Australian Standard*.

Department of Planning (1990). Risk criteria for land use safety planning. *Hazardous Industry Planning Advisory Paper no. 4*. Department of Planning, Sydney.

European Railway Agency (2016). Assessment of achievement of safety targets. *Report*.

Health and Safety Executive (UK) (1988). The tolerability of risk from nuclear power stations. London, HMSO.

Higson, D. J. (1989). Risks to individuals in NSW and in Australia as a whole. *ANSTO report*.

IEC 62290 (2014). Railway applications - Urban guided transport management and command/control systems

ICAO (2015). Manual on Remotely Piloted Aircraft Systems (RPAS).

NSW Department of Planning. (2011). Hazardous Industry Planning Advisory Paper No 4 – Risk criteria for land use safety planning (HIPAP-4).

HSE-INDG271 (2011). Buying new machinery – A short guide to the law and your responsibilities when buying new machinery for use at work.

IEC 62290-1.2014 (2014). Railway applications - Urban guided transport management and command/control systems. International Standard.

JARUS. (2015) Scoring paper to AMC RPAS.1309 Issue 2. WG-6 Report.

Lees, F. (2011). Loss prevention in the process industries – Third Edition. *Elsevier*.

Office of the National Rail Safety Regulator (ONRSR) (2012). National Rail Safety Law.

Ward, D. (2011). System safety in hybrid and electric vehicles. Australian System Safety Conference 2011.

Additional readings:

Aerospace – systems engineering and space – SAE ARP 4761 (1996. Guidelines and methods for conducting the safety assessment process on civil airborne systems and equipment

Defence post-war reliability – British Standards Institution, 1994. Reliability of systems, equipment and components

Energy /Power – Three Mile Island 1979 – Perrow. C (1984) Normal accidents

Hazardous Chemicals – Lees, F (1980) – Loss prevention in the process industries

Rail – Commission of Inquiry into the Waterfall Rail Accident, 2004, National Rail Safety Law and Regulations

Safety Critical – AS 61508 Functional safety of electrical /electronic /programmable electronic safety related systems