Phase 2 of the Economic Active Transport Project to deliver a best practice method to cost the health benefits of active transport in NSW

Preliminary report

Version March 2022

Mary Wanjau[1], Holger Möller[2,3], Fiona Haigh[4,5], Belen Zapata-Diomedi[6], Lennert Veerman[1]

[1] Griffith University, School of Medicine

[2] The George Institute for Global Health

[3] School of Population Health and Community Medicine, UNSW Sydney

[4] Health Equity Research Development Unit (HERDU), UNSW Sydney

[5] Sydney Local Health District

[6] RMIT Centre for Urban Research

Table of Contents

List of Tables

List of Figures

List of abbreviations

ABS	Australian Bureau of Statistics
ACM	All-cause mortality
ADS	Australian Demographic Statistics
AIHW	Australian Institute of Health and Welfare
AMSTAR	Assessment of Multiple Systematic Reviews
AQI	Air Quality Index
ATAP	Australian Transport Assessment and Planning
BMI	Body Mass Index
CASP	Critical Appraisal Skills Programme
CESD	Center for Epidemiologic Studies Depression Scale
CI	Confidence Interval
COPD	Chronic obstructive pulmonary disease
DSM	Diagnostic and Statistical Manual of Mental Disorders
DW	Disability weight
GBD	Global Burden of Disease
GRADE	Grading of Recommendations, Assessment, Development, and Evaluation
HALYs	Health adjusted life years
HR	Hazard ratio
IHME	Institute for Health Metrics and Evaluation
ITHIM	Integrated Transport and Health Impact model
LBP	Low Back Pain
LTPA	Leisure time physical activity
MeSH	Medical subject heading search
MET	Metabolic equivalent of task
MI	Myocardial infarction
MSLT	Multi-state life table
MVPA	Moderate-to-vigorous-intensity physical activity
NCDs	Non-communicable diseases
NHS	National Health Survey
NSW	New South Wales
OA	Osteoarthritis
OBPR	Office of Best Practice Regulation
OR	Odds ratio
PA	Physical Activity
PHS	Population Health Survey
PIF	Potential Impact Fraction
$PM_{2.5}$	Particulate Matter 2.5
PRISMA-P	Preferred Reporting Items for Systematic reviews and Meta-Analysis Protocols
RR	Relative risk
SD	Standard deviation
TfNSW	Transport for New South Wales
VSL	Value of statistical life
VSLY	Value of a Statistical Life Year

WCRF	World Cancer Research Fund
WHO	World Health Organisation
YLD	Years lived with disability

Executive Summary

Physical inactivity is one of the main contributors to the rise in non-communicable diseases worldwide. Active transport, which refers to modes of transport that involve physical activity such as walking and cycling, and walking and cycling to and from public transport transit stops, offers a promising means of raising physical activity at population level. The health benefits of active transport are well established but there is no agreed method of valuing them in strategic business cases in New South Wales to date. The aim of this project was to deliver a best practice method to cost the health benefits of active transport in NSW.

Quantifying the health-related economic benefits of active transport can be considered as a three-step process:

1. Estimation of the impact of infrastructure measures on active transport behaviour and impact of changed transport behaviour on exposures related to active transport
2. Quantification of the impact of changes in exposures on health outcomes
3. Costing the health benefits

This project focused mainly on steps 2 and 3. For the 1st step, the input for the model is the net change in travel based on the assumption that prior traffic/travel modelling has been able to estimate the net change in travel-related physical activity. It leaves open the possibility that, for example, on a given stretch of new infrastructure, some active travellers have not changed mode but simply diverted their travel. The model currently assumes that new active travellers shifted away from car travel.

The preceding Phase 1 project identified the 'Zapata-Diomedi model' as current best practice but recommended to use NSW-specific data and to consider additional health outcomes associated with physical activity and air pollution, not considered in the 'Zapata-Diomedi model' for inclusion in the model.

With oversight from the cross-agency advisory group, in the current Phase 2 project, we created the *NSW Active Transport Health Model*. Building on the Zapata-Diomedi model, it uses a multistate lifetable model to quantify the health impacts of changes in active transport behaviour. The multistate lifetable method models the impact of active transport over the lifetime of a population and allows for different health states and co-morbidities.

The updated model considers exposures associated with active transport: physical activity, air pollution and road trauma. Health outcomes associated with physical activity are breast cancer, colon cancer, type 2 diabetes, ischemic stroke and ischemic heart disease. Additional health outcomes associated with physical activity that were investigated for inclusion in the model were depression, anxiety, osteoarthritis, low back pain, and all-cause mortality. For air pollution the model includes ischemic stroke, ischemic heart disease, tracheal, bronchus and lung cancer, chronic obstructive pulmonary disease, lower respiratory tract infection, type 2 diabetes, intracerebral haemorrhage, and subarachnoid haemorrhage. For road trauma, mortality and morbidity by mode of transport is included.

Systematic reviews of peer-reviewed evidence were carried out to assess the strength of the evidence for the association between physical activity and the health outcomes: all-cause mortality, musculoskeletal diseases (low back pain, osteoarthritis), mental health problems (depressive

disorders, anxiety). Criteria for causal interpretation were also applied. The evidence that physical activity can prevent depression and anxiety was judged to be 'probable'. The evidence that physical activity can prevent premature mortality was considered convincing. These have been included in the model. However, for osteoarthritis and low back pain, the evidence was 'possible', and these two conditions are suggested to be used in sensitivity analysis only.

The *NSW Active Transport Health Model* produces a range of outcomes including health adjusted life years (HALYs), life years lived, years lived with disability (YLDs), health care costs, incidence, prevalence and mortality (numbers and percentage change) for the included health conditions over the first 25 years following the start of the intervention and over the lifetime. HALYs are valued in monetary terms (three options are modelled) and values per additional km of walking and cycling are reported. The interface ('dashboard') allows the analyst a choice of a range of options with respect to scenarios modelled and the in- or exclusion of risk factors and health conditions.

The central (reference case) scenario was based on the following: an intervention in which 10,000 travellers replace 5 weekly car trips of 0-2km with walking over 1 year; added walking is modelled as proportional increase in walking by age and sex groups; all levels of physical activity within each age-sex group benefit equally; ages 20-100 years; costs discounted at 7% and health at 3% per annum; effects via exposures to physical activity, road trauma, and air pollution (ambient and traffic participants); 100% of active travel-related physical activity is net additional physical activity; all health conditions discussed above are included except osteoarthritis and low back pain; inclusion of all-cause mortality based on studies in which physical activity is measured objectively with accelerometry; HALYs are valued in accordance with current *Office of Best Practice Regulation* recommendations.

The reference case puts the economic value of the health benefits of a km walking at $5.42 (95% uncertainty interval $5.11 to $5.78). An additional km cycled results in $1.47 ($1.38 - $1.58) in health benefits. Assuming no risk of injury and background levels of air quality, an additional km of cycling on separate, off road bikeways is valued at $1.58 ($1.48 to $1.69) per km. An increase in walking associated with public transport use is associated with health benefits of $3.53 ($3.29 to $3.79) per km.

These values are sensitive to a range of factors, including:

- age ($0.90 net benefit for ages 20-29, 30-fold higher benefits for 70+),
- the risk factors included (large benefits for physical activity, small benefit for ambient air quality, small negative effect of air quality for traffic participants, moderate effects either way of road trauma),
- the assumption that all physical activity is additional (a rapid systematic review of published evidence [Appendix P] shows that active travel-related physical activity does not displace physical activity in other domains of life),
- the diseases included (the addition of mental health doubles the estimates when all-cause mortality is switched off),
- the direct effect of physical activity on mortality,
- the choice of the evidence base for that effect (the use of evidence from studies using accelerometry measured physical activity to estimate the effect on all-cause mortality increases the estimated value by a factor 8.7 compared to not including all-cause mortality),
- the size of the intervention (small interventions give higher estimates than larger interventions because of diminishing returns to physical activity),
- the duration of the intervention,

- and indirectly, the nature of the intervention (e.g., because cycling is mainly done by young people, additional trips are allocated to that age group too, where the benefits are lower than for older groups),
- the choice of the value for a HALY matters; the results using the ATAP (Australian Transport Assessment and Planning)/TfNSW value are almost 70% higher than those using the value recommended by the Commonwealth Office for Best Practice Regulation, and
- the rate at which future health gains are discounted.

Compared to previous estimates, the health benefits per km walking and cycling found in this study are higher for walking and consistent for cycling. Both per km walking and cycling values are considerably higher when using a lower discount rate for health outcomes. with previous estimates. A literature review for the TfNSW Economic Parameter Values reported values per km walking between $0.44 and $2.44, and between $0.07 and $1.31 for cycling [1]. The higher values in this study are mainly due to the inclusion of additional outcomes for physical activity – all-cause mortality and mental health problems.

The proposed method conforms to Treasury and Transport for NSW guidelines, other than that in the reference case, health is discounted at the social discount rate of 3% rather than the 7% recommended for the opportunity cost of capital [2]. The method can be used at two levels. Firstly, the per-km values can be used in cost benefit analyses for business cases for infrastructure projects. Secondly, the model can be used to run analyses with the parameters set by the analyst to align with the characteristics of the project. The broader methodology can be used for *health impact assessment* in other policy areas where interventions outside the health sector have health consequences. Additional steps to account for non-health benefits are required for a full cost-benefit analysis.

1. The creation of the NSW Active Transport Health Model

1.1 Background

Physical inactivity is one of the main contributors to the rise in non-communicable diseases (NCD) such as diabetes, heart disease and cancer worldwide [3, 4]. Active transport, which refers to modes of transport that involve physical activity such as walking and cycling, and walking and cycling to and from public transport transit stops, is increasingly recognised as a very promising means of enhancing physical activity (PA) at population level, thereby reducing the burden of NCDs [5, 6]. Although the health benefits of active transport have been highlighted [7-12], there is no formally recognised method of valuing the health benefits of active transport in strategic business cases in New South Wales (NSW) to date.

Against this background, the NSW Ministry of Health initiated a project (Phase 1) to identify a best practice method to cost the health benefits of active transport. The project identified a three-step process to quantify the health-related economic benefits of active transport (

Figure 1):

1. Estimation of the impact of infrastructure measures on active transport behaviour and impact of changed transport behaviour on exposures related to active transport
2. Quantification of the impact of changes in exposures on health outcomes
3. Costing the health benefits

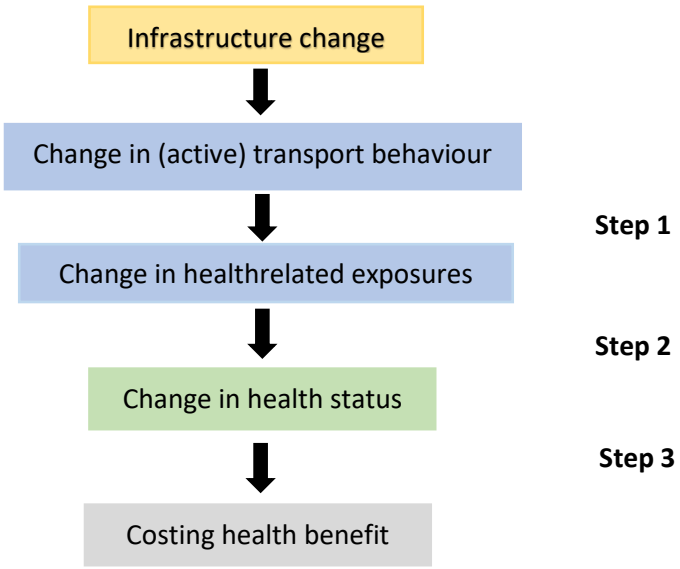


Figure 1 Three-step process of costing the health benefits of active transport

Based on a systematic review of the international evidence and evaluation of the models against a set of criteria identified in stakeholder consultation (Appendix A), a proportional multi-state life table (MSLT) model, the '*Zapata-Diomedi model',* was recommended as the best practice model to quantify the impact of active transport on health outcomes [7]. This model has previously been used to quantify the health benefits of a shift from motorised to active transport in Brisbane Queensland, Australia [7]. The model considered all risk factors identified in the systematic review that are associated with active transport: PA, air pollution and road transport injury (Figure 2). Health outcomes for PA were breast cancer, colon cancer, diabetes mellitus, ischemic stroke, and ischemic heart disease (IHD). These were the conditions included in the latest update of the Global Burden of Disease (GBD) study available at the time [13]. For air pollution, Zapata-Diomedi and colleagues included the health outcomes ischemic stroke, IHD, tracheal, bronchus and lung cancer, and chronic obstructive pulmonary disease (COPD). These were based on World Health Organisation (WHO) methods and tools for assessing the health risks of air pollution [14] and a systematic review of outdoor particulate matter exposure and lung cancer [15]. Road transport injuries included in the Zapata-Diomedi model were fatalities and injuries by road user type and striking mode. Outcome measures in the model were health care costs, life years, health adjusted life years (HALYs), prevalent cases, deaths and years lived with disability (YLDs).

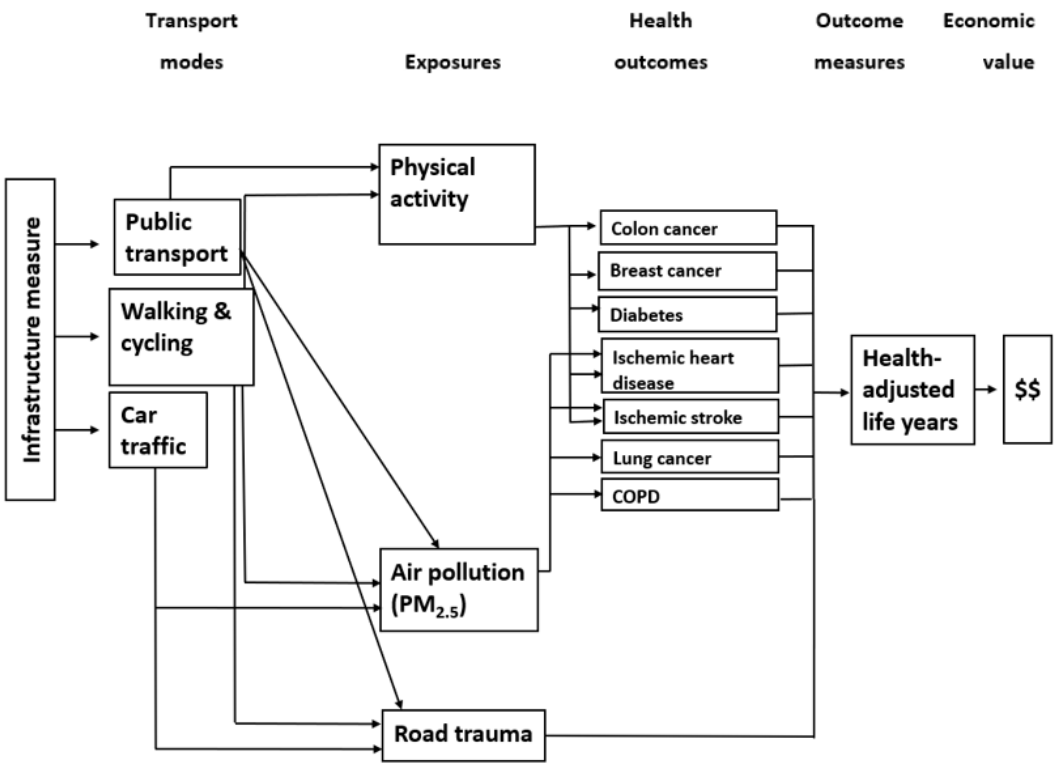


Figure 2 Schematic overview of the Zapata-Diomedi model

In the evidence review and during stakeholder consultation in Phase 1 of the project it emerged that, in addition to the health outcomes currently considered in the Zapata-Diomedi model [7], further health outcomes associated with PA and air pollution should be considered for inclusion in the model. These were mental health problems, musculoskeletal diseases and all-cause mortality (ACM)

associated with PA, and respiratory tract infection, type 2 diabetes, intracerebral haemorrhage and subarachnoid haemorrhage associated with air pollution. Moreover, it was recommended to populate the model with NSW data and to review if the underlying model assumptions needed updating. These model extensions and updates were beyond the scope of Phase 1 of the project. Consequently, the NSW Ministry of Health commissioned the project team to address these issues in Phase 2 of the project.

Phase 2 of the project aimed to create the 'NSW Active Transport Health Model'. This model was derived from the best practice model identified in Phase 1 [7]. In phase 2, the model was populated with NSW data, model assumptions were updated where new epidemiological evidence was available, and the model was extended to include additional health outcomes associated with PA and air pollution.

Towards this aim of delivering a best practice method to cost the health benefits of active transport in NSW, Phase 2 progressed in four main phases:

1. Creation of the *NSW Active Transport Health Model* by updating the Zapata-Diomedi model for the NSW context using NSW specific data sources and updating underlying assumptions
2. Extending the model to provide for additional health outcomes for PA and air pollution
3. Reviewing and documenting the current methods used by Transport for NSW (TfNSW) to provide the context of current practice in NSW
4. Preparing the NSW Active Transport Health Model for use in NSW practice

All the project phases have been carried out in close collaboration with the cross-agency advisory group consisting of representatives from six NSW government clusters: Treasury; Transport; Planning, Industry and Environment; Premier and Cabinet; Education; and Health.

1.2 NSW Active Transport Health Model

The *NSW Active Transport Health Model* estimates the impact of a change in active transport behaviour by simulating the population over their remaining lifetime, once under 'business as usual' assumptions, and in parallel, in a scenario in which more people engage in active transport. The proportional MSLT allows for the inclusion of multiple diseases whilst allowing for comorbidities. The impact is calculated by taking the difference in outcomes between these two scenarios.

A schematic description of the proportional MSLT is presented in Figure 3. In this project, a direct impact of PA on mortality from any cause was also modelled in some of the analyses. Additional detail of the disease model is presented in Appendix B.

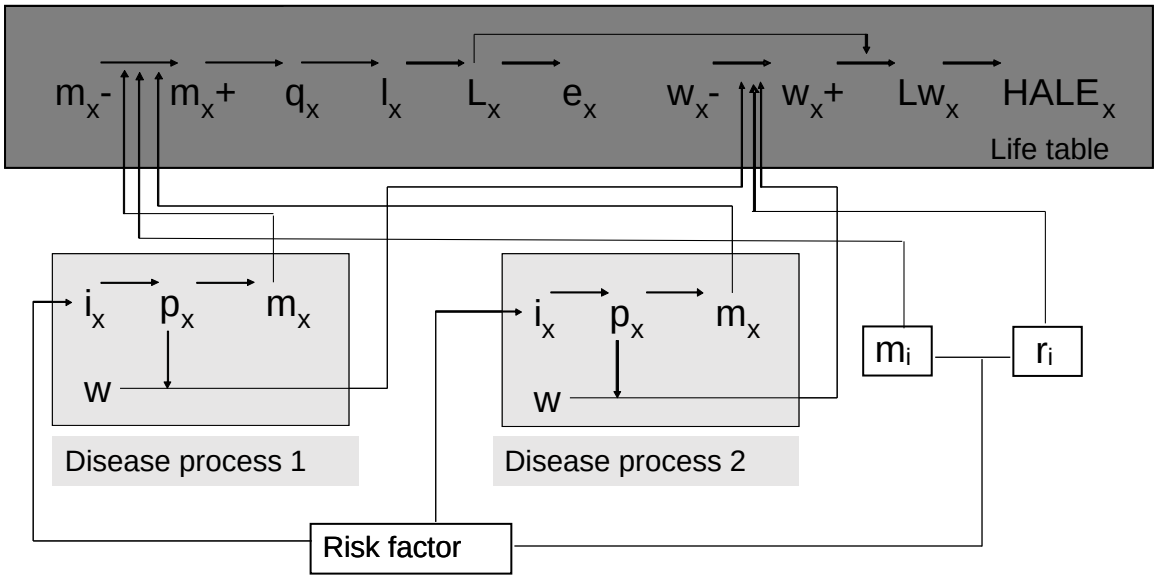


Figure 3 Figure Schematic description of a proportional MSLT

This figure describes the interaction between life table parameters and disease parameters. All the parameters are age-specific denoted with x, i is incidence, p is prevalence and m is mortality, w is disability adjustment, q is probability of dying, l is number of survivors, L is life years, Lw is disability adjusted life years and HALE is health adjusted life expectancy, '-'denotes parameter related to diseases or causes that specifically excludes modelled diseases or injuries and '+' relates to all modelled diseases in the model. A change in the determinant of health (physical activity and $PM_{2.5}$) translates into changes in incidence (i_x), which changes disease specific prevalence (p_x) and mortality (m_x). For presentation purposes we only depict two diseases processes, however, in this study we modelled 14 diseases (breast cancer, colon cancer, ischemic stroke, type 2 diabetes, ischemic heart disease, depression, anxiety, osteoarthritis, low back pain, tracheal, bronchus and lung cancer, chronic obstructive pulmonary disease, respiratory tract infection, intracerebral haemorrhage and subarachnoid haemorrhage). Road fatalities (m_i) impact directly on mortality. Road injuries (r_i) impact on years lived with disability, which are captured in Lw in the schematic description.

1.3 Update of the Zapata-Diomedi model for the NSW context

The Zapata-Diomedi model was reviewed to identify:

- the model input parameters that needed updating for the model to be NSW specific
- the parameters that needed updating because new evidence is available

We found that the epidemiological data needed to be updated to reflect new evidence. New estimates for health care costs were also available (see section 1.5). Various NSW data sources identified in Phase 1 of the project (Appendix C) were accessed and checked for suitability for use in the model. With input and consultation from the Ministry of Health, TfNSW and the project advisory group, various data sources were identified for use in the updated model. The rationale for each choice is presented in Table 1.

Table 1 Selected data sources for model input parameters and rationale

Data	Source	Rationale
Population	Australian Bureau of Statistics. 3101.0 - Australian Demographic Statistics, Jun 2019	Most up to date data and representative of the NSW population.
Mortality rates Years lived with disability (YLD) total rate Disease incidence Disease prevalence Disease remission Disease case fatality Disability weights	GBD 2017 [16, 17]	Most up to date, comprehensive epidemiological evidence
Prevalence data on PA for NSW	2017-2018 National Health Survey CURF data - Australian Bureau of Statistics [18]	Only survey to provide information on total PA including PA at the workplace (not including travel to work).
Air pollution data	NSW Planning, industry and environment Air Quality Index (AQI) data [19]	Provides data on $PM_{2.5}$ from monitoring stations in NSW
NSW road transport injury data	TfNSW Centre for Road Safety Crashlink database [20]	Provides information on all police reported crashes in NSW
Health care costs	AIHW Disease Expenditure in Australia 2015-16 [21]	Most recent comprehensive set of estimates by health condition

1.4 Extension of the Zapata-Diomedi model to provide for additional health outcomes for physical activity and air pollution

In Phase 1 of the project, additional health outcomes associated with PA and air pollution were identified and recommended for inclusion in the NSW Active Transport Health Model. These outcomes were depression, anxiety, osteoarthritis (OA), low back pain (LBP) and all-cause mortality (ACM) for the PA exposure and lower respiratory tract infection, type 2 diabetes, intracerebral and subarachnoid haemorrhage for air pollution (

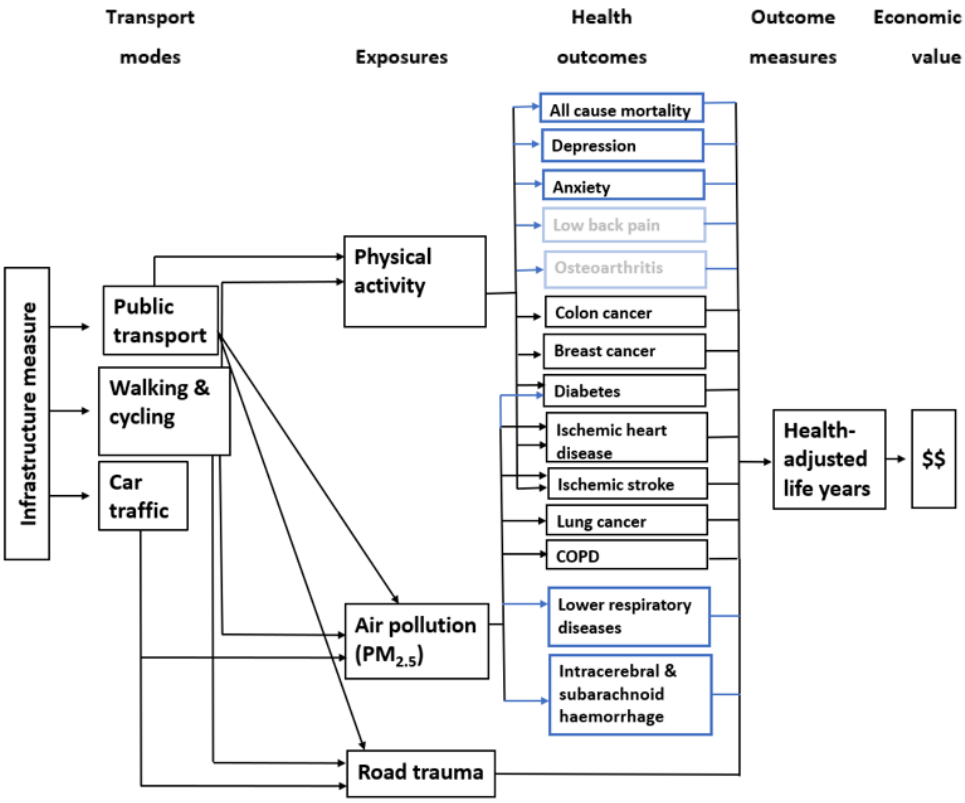


Figure 4). In this project phase we reviewed the epidemiological evidence on the strength of the association between these outcomes and exposure to PA and air pollution and assessed the likelihood of these associations being causal. The identification of these additional health outcomes was informed by the evidence gathered in Phase 1 of the project and by the work of the GBD study international expert working groups [16].

In the Zapata-Diomedi model, the following health outcomes were considered to be causally associated with PA: breast cancer, colon cancer, type 2 diabetes, ischemic stroke, and ischemic heart disease. These outcomes are also considered in the latest version of the GBD study [16]. Since inclusion in the GBD study is preceded by a thorough evaluation process, we considered this as sufficient justification for inclusion of these health outcomes in the NSW Active Transport Health Model. As described in the methods appendix to the GBD 2017 study [22], only the risk- outcome pairs that were assessed as meeting the World Cancer Research Fund (WCRF) grades of 'convincing' or 'probable' evidence were included in the study [23].

The additional health outcomes associated with PA (depression, anxiety, OA, LBP and ACM) suggested for inclusion in the updated model, were not considered in the 2017 GBD study. Hence, we carried out three systematic reviews of the latest epidemiological evidence to make a judgment on the strength of the evidence for the causality of the observed associations between these outcomes and PA (see section 4).

For air pollution exposure, the health outcomes included in the Zapata-Diomedi model were ischemic stroke, IHD, tracheal, bronchus and lung cancer, and COPD. Suggested additional health outcomes for air pollution exposure were lower respiratory tract infection, type 2 diabetes, intracerebral haemorrhage and subarachnoid haemorrhage. In the latest update of the GBD study [16], these four additional outcomes have been included as outcomes of air pollution exposure. As mentioned earlier, since inclusion in the GBD is preceded by a thorough evaluation process, we considered this as sufficient justification for inclusion of these health outcomes in the NSW Active Transport Health Model. Health outcomes associated with road trauma in the model remained road transport injuries and deaths, by mode of transport. The extensions to the updated model are illustrated below in an analytical framework for modelling the health impact of active transport (

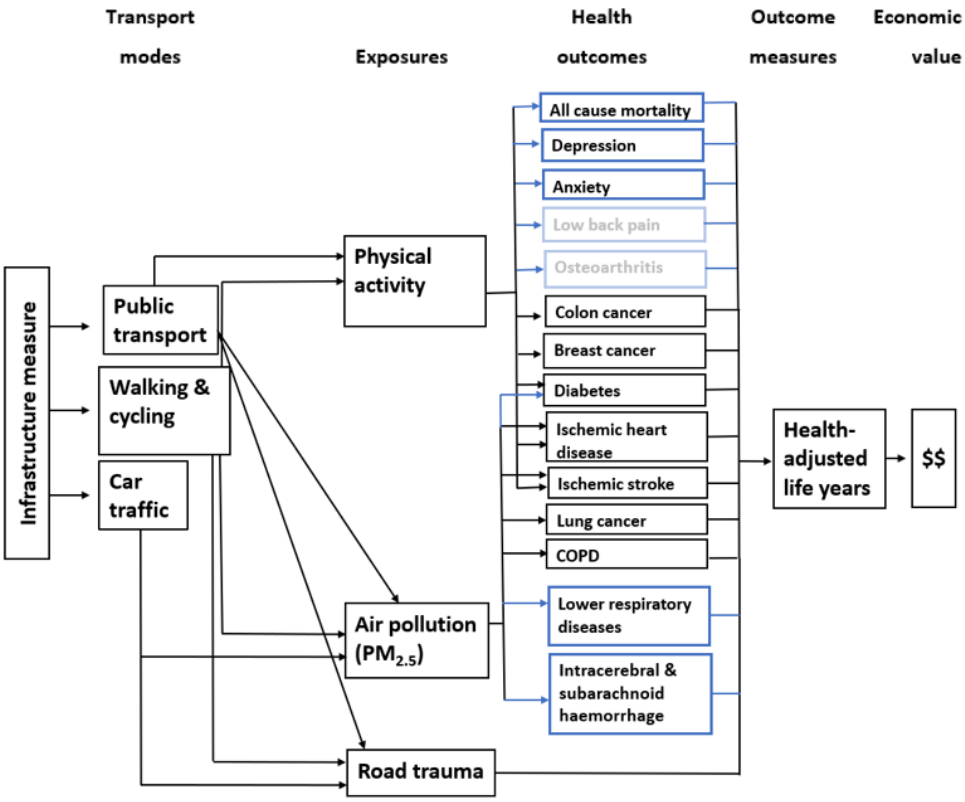


Figure 4).

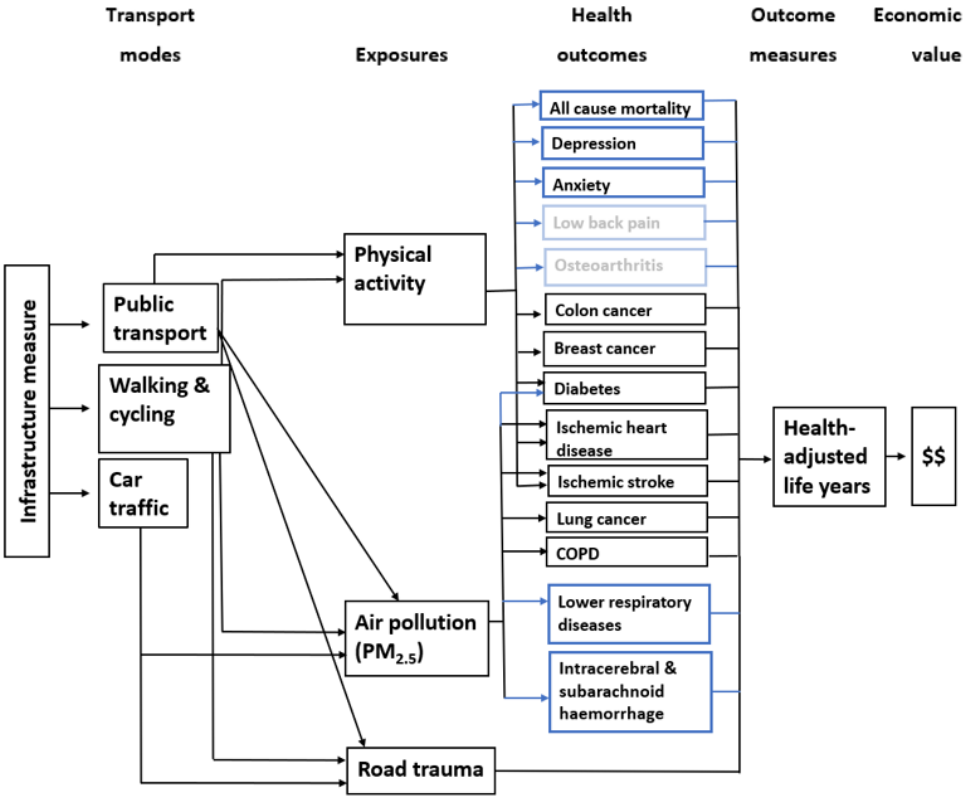


Figure 4 Analytical framework for modelling the health- and health related economic impacts of active transport

Infrastructure changes that consider active transport affect public transport use, car use and active transport behaviour. Increased active transport results in increased physical activity and exposure to on-road air pollution, noise, and risk of road trauma. Decreased car use impacts on road trauma and air pollution. Increased public transport use promotes physical activity though walking to public transport stations and impacts on air pollution. The effect of exposures on health outcomes listed is modelled and HALYs gained can be calculated. Cost benefits are calculated by multiplying HALYs with the value of a year in full health. Blue arrows and boxes indicate additional health outcomes considered for the NSW Active Transport Health Model. Outcomes in grey (low back pain and osteoarthritis) are included in the model but are not included in the reference case analysis.

1.5 Preparation of epidemiological input data

To populate the model, incidence, prevalence and cause specific mortality data from the GBD 2017 Results tool were used [17]. Prevalent YLD rates (which indicate the average loss of quality of life due to disease, by age) were also retrieved from the GBD 2017 data. In addition, the NSW population data were retrieved from the Australian Bureau of Statistics [24], and the healthcare costs were calculated using data from the Australian Institute of Health and Welfare (AIHW) [21].

We used specialised software (DisMod II) to enforce internal consistency in the epidemiological estimates, and to derive parameters not provided in the data sources [25]. More detail on this is presented in Appendix D.

For health care costs, we used data from the *Disease Expenditure in Australia 2015-16* report prepared by the Australian Institute of Health and Welfare [21]. The report provides a comprehensive set of estimates of health expenditure by conditions in disease groups, demographics, and areas of expenditure.

For the calculations of costs per case, GBD 2017 disease prevalence, incidence, YLD, and population data were used [17]. Detailed costs for male, female for all age groups are presented Appendix E and Appendix F.

In addition to costs for the specific health conditions in the model, overall health care costs for all other health conditions are also included. This is necessary because as interventions prolong life, health care costs will be incurred in these 'health care costs in added years of life' [26]. See Appendix F.

We inflated the costs to the June 2019 values. We used an inflation factor that we calculated from the Total health price index and industry-wide indexes, 2007–08 to 2017–18, extrapolating to 2019. The total health price index was sourced from AIHW health expenditure database [27].

2. Exposures

The model includes the exposures related to active transport identified for inclusion in the model in Phase 1 of the project: PA, road trauma and air pollution ($PM_{2.5}$).

2.1 Physical activity

We used PA data for NSW from the National Health survey (NHS) 2017-2018 [18]. The NHS was preferred to the NSW Population Health Survey (PHS) [28] because it includes information on total physical activity, inclusive of workplace physical activity. Workplace physical activity was measured as the amount of physical activity at work on a typical workday (not including travel to and from work). In comparison, the NSW PHS gives data on activity for recreation or exercise or to get to or from places and does not give data on physical activity at work. Weekly physical activity levels were calculated by multiplying the minutes of total weekly physical activity by the average metabolic equivalent of task (MET) value for each activity (walking= 3.5, cycling= 6.8, moderate PA=5, vigorous PA= 7.5). The MET value estimates were selected from the Compendium of physical activities [29]. A MET is defined as the ratio of work metabolic rate to a standard resting metabolic rate of 1.0. As 1 MET is considered the resting metabolic rate obtained during quiet sitting, in the Compendium, activities are listed as multiples of the resting MET level. This ranges from 0.9 METs (sleeping) to 18 METs (running at 17.5 km/hr). In the model, we subtracted 1 from the MET value for each activity, on the assumption that the activity displaces sitting with a MET value of 1.

For the model, PA levels were categorized into the four groups: inactive, low active (>0 & <600 MET-minutes per week), moderately active (>=600 & <1600 MET-min/wk) and highly active (>=1600 MET-min/wk) (Table 2) [18].

Table 2 Work, transport and exercise activity categories NSW 2017-2018

	Men				Women			
Age group	Inactive	Low active	Moderately active	Highly active	Inactive	Low active	Moderately active	Highly active
20 - 24	0.0%	11.3%	20.8%	67.9%	2.6%	19.2%	32.1%	46.2%
25 - 29	7.4%	13.0%	20.4%	59.3%	10.5%	18.9%	28.4%	42.1%
30 - 34	8.1%	17.1%	24.3%	50.5%	9.7%	28.4%	29.0%	32.9%
35 - 39	9.9%	21.8%	21.1%	47.2%	12.1%	27.9%	33.3%	26.7%
40 - 44	11.4%	17.9%	20.3%	50.4%	14.7%	39.0%	22.1%	24.3%
45 - 49	10.0%	20.0%	22.9%	47.1%	12.4%	27.7%	27.0%	32.8%
50 - 54	15.7%	24.5%	18.6%	41.2%	10.1%	32.9%	25.3%	31.6%
55 - 59	12.1%	24.2%	27.3%	36.4%	17.5%	25.2%	31.5%	25.9%
60 - 64	20.6%	21.3%	25.0%	33.1%	17.3%	32.1%	29.5%	21.2%
65 - 69	22.7%	27.3%	28.9%	21.1%	26.5%	23.8%	25.2%	24.5%
70 - 74	17.3%	30.6%	25.5%	26.5%	17.7%	41.5%	24.5%	16.3%
75 - 79	26.9%	25.6%	32.1%	15.4%	23.5%	35.3%	28.4%	12.7%
80 - 84	28.6%	26.5%	28.6%	16.3%	30.8%	49.2%	9.2%	10.8%
85+	31.6%	34.2%	18.4%	15.8%	49.3%	37.7%	8.7%	4.3%

Inactive (0 MET), low active (>0 & <600 MET-min/wk), moderately active (>=600 & <1600 MET-min/wk) and highly active (>=1600 MET-min/wk)

Source of data inputs: Australian Bureau of Statistics. Microdata: National Health Survey, 2017-18 [18].

The mean minutes and kilometres walking for those using public transport varied slightly by age and gender (Table 3). For the model, this means that a mode shift towards greater use of public transport benefits all ages and genders.

Table 3 Mean minutes and kilometres walking for those using public transport

Age	Gender	Minutes	Km
17-49	Male	13.69	1.02
17-49	Female	13.15	0.98
50-74	Male	12.58	0.94
50-74	Female	12.77	0.95
75 plus	Male	10.03	0.75
75 plus	Female	13.31	0.99

This was based on analysis of data from the Australian Health Survey: Physical Activity, 2011±12 [Internet] 2015, as described in Zapata-Diomedi, Knibbs [7].

For the analysis of all-cause mortality, the data were also analysed by quartiles of activity (Table 4). This aligns with the values reported in the systematic review by Ekelund and colleagues, who pooled results from cohort studies that used objectively measured PA (with accelerometers) as exposure [30]. To use their results, we had to assume that the activity patterns in NSW are similar to those in the source studies. This seems a reasonable assumption. If anything, participants in cohort studies will tend to be healthier and more active than the general population, which could result in underestimation of the mortality risk among the least active members of society. From the process, we derived a risk function (a level of risk at every level of PA), which was then applied to the activity categories used in the model.

Table 4 Mean work, transport and exercise activity MET for activity quartiles, men Australia 2017-2018

	Men	Women
	Mean (95% CI)	Mean (95% CI)
Least active	67.0 (58.0-76.1)	30.4 (26.0-34.8)
Second quartile	659.7 (636.1-683.4)	346.4 (334.7-358.0)
Third quartile	1871.6 (1818.8-1924.3)	1048.3 (1022.0-1074.6)
Most active	8723.3 (8142.9-9303.7)	4602.2 (4228.0-4976.5)

Source ABS National Health Survey [18]
The active transport modes included in the model are walking, cycling, and walking and cycling to and from public transport transit stops.

2.2 Air Pollution

Particulate matter ($PM_{2.5}$) was used as a general indicator of air pollution. $PM_{2.5}$ are very small particles with a diameter of 2.5 micrometres or less found in smoke and other air pollutants. $PM_{2.5}$ is commonly used as an indicator of air quality and measured by air quality monitoring sites. As representative data for NSW were not available, data on $PM_{2.5}$ from air quality monitoring sites measuring $PM_{2.5}$ in the greater Sydney area for 2018 were obtained from the NSW Planning Industry and Environment website [19]. Annual averages were calculated. Days with $PM_{2.5}$ equal to or above the 95th percentile were excluded from the analysis as these higher air pollution levels are typically due to landscape fires [31]. After excluding days with potential landscape fires, the mean annual $PM_{2.5}$ level for all 14 stations was 7.11 µg/m^3 [95% CI 7.07-7.16]. The $PM_{2.5}$ pollution attributable to road transport was calculated from a NSW modelling study which estimated that around 7.7% of ambient $PM_{2.5}$ in the greater Sydney region was estimated to be due to on-road vehicles [32].

The model incorporated the health effects of exposure to $PM_{2.5}$ via two mechanisms: population-wide ambient effects and individual effects among traffic participants. Change in exposure for the ambient effect was estimated as the difference in the mean $PM_{2.5}$ between baseline and intervention scenario, adding up the emissions due to the various traffic modes. Based on previous studies [33-35], the change in individual exposure due to increase in active transport was calculated as follows:

$$\Delta\,\textbf{individual exposure} = \big((\textbf{Total dose active transport scenario}/\textbf{Total dose baseline}) - \mathbf{1}\big) * \boldsymbol{background\ exposure}$$

Equation A: Change in individual exposure to $PM_{2.5}$

To estimate the total $PM_{2.5}$ dose per week for the baseline (status quo) and active transport scenario, activity-specific information was needed for: time allocation in hours per week, ventilation rates, concentration of $PM_{2.5}$ and total inhaled dose of $PM_{2.5}$. The only difference between baseline and intervention scenarios is that part of the car occupancy time at baseline is replaced by active transport for the intervention scenario.

Using ventilation rates per minute and concentration of $PM_{2.5}$ for each of the activites previously applied in a similar study [35] we calculated the total inhaled dose per week for each activity. Sleep and resting time have the equivalent to the background $PM_{2.5}$ concentration, whereas all other activities have higher concentration levels.

2.3 Road trauma

In New South Wales, according to the Road Transport Act 1999, a crash must be reported to police when any person is killed or injured, when there is damage of over $500 to property other than the vehicles concerned, when drivers involved in the crash do not exchange insurance and contact information, when one or more of the drivers is reported to be driving under the influence of alcohol, or if a vehicle involved in the crash is towed away. In this study, all road transport crashes that were recorded by NSW police and resulted in injury or deaths were included in the analysis.

A distance-based model was used to estimate the effect of change in active transport behaviour on road trauma. Injuries and fatalities for transport modes were summarised by combinations of victim and striking mode, using 2018 data from TfNSW Crashlink database (Appendix G). Road crashes involving only one transport mode were presented as “single mode”.

Baseline injury and fatality rates were calculated per pairwise combination of victim and striking mode [8, 36]. Calculating rates per pairwise combination requires data on the number of injuries/fatalities, person-kilometres travelled for the victim mode, and vehicle-kilometres travelled for the striking mode per year (Equation B). Person-kilometres travelled were derived from the NSW household travel survey, the Survey of Motor Vehicle Use 2018 [37] and the NSW cycle survey [38].

For pedestrians, bicyclists and motorcycle/moped, person-kilometres travelled were the same as vehicle kilometres travelled. The number of injuries and fatalities for change in the transport mode mix (Appendix G) were estimated using Equation C. These calculations take into consideration a safety in numbers effect whereby increases in pedestrian and cyclist numbers results in a less than proportional increase in the number of injuries [39].

$$R_0 = \frac{Number\ of\ injuries_{Victim}}{(PKM_{Victim} * VKT_{Striking\ vehicle})^{0.4}}$$

Equation B Rate of injuries per pairwise combination of victim and striking modes

$$Number\ of\ injuries_{Victim_0} = R_0 * (PKM_{Victim} * VKT_{Striking\ vehicle})^{0.4}$$

Equation C Number of injuries per pairwise combination of victim and striking modes

3. Updated measures of disease association

3.1 Physical activity

For the health outcomes already included in the Zapata-Diomedi model, evidence on the association between PA and breast cancer, colon cancer, type 2 diabetes, ischemic stroke and IHD was assumed sufficiently strong to warrant inclusion of these conditions in the model. Internationally, the GBD study regularly quantifies the burden of disease attributable to risk factors such as PA and air pollution among others [40]. As part of this, an international expert working group reviews the evidence on these risk factors annually. We used the evidence gathered in the GBD studies to inform selection of the measures of strength of association between exposure and the health outcomes in our model that were included in the GBD studies (Breast Cancer, Colorectal Cancer, Diabetes, Ischaemic Heart Disease, Ischemic Stroke). The estimates of the measures of disease association [41, 42] in the GBD 2013 [43] and GBD 2017 [16] studies were reviewed. As with the Zapata-Diomedi model, the measures of association reported in the GBD 2013 were selected for use in the updated model. The measures of association used in the GBD 2017 were taken from a systematic review and dose-response meta-analysis study done by Kyu and colleagues [42]. This study was not used in the model because the authors categorised physical activity in very high levels of MET minutes/week (<600, 600-3999, 4000-7999 and, ≥8000). The selected measures of association were based on the study by Danaei and colleagues [41] that aimed to estimate the mortality effects of 12 modifiable risk factors in the United states. In this comparative risk assessment, the authors obtained the relative risk for each PA exposure category for the diseases with probable or convincing causal associations, based on epidemiological evidence from systematic reviews and meta-analyses. The magnitudes of relative risk for the effects of PA presented in this study were applied in the updated model (Table 5).

Table 5 Magnitudes of relative risks for the effects of physical inactivity on disease-specific mortality

Outcome (mortality)	Age (years)	Inactive	Insufficiently Active	Recommended Level Active	Highly Active
Breast Cancer	30-44	1.56 (1.30 - 1.87)	1.41 (0.84 - 2.36)	1.25 (0.90 - 1.74)	1
Females	45-69	1.67 (1.44 - 1.94)	1.41 (0.84 - 2.36)	1.25 (0.90 - 1.74)	1
	70-79	1.56 (1.32 - 1.84)	1.36 (0.72 - 2.57)	1.25 (0.79 - 1.98)	1
	80+	1.45 (1.16 - 1.81)	1.32 (0.55 - 3.19)	1.25 (0.62 - 2.51)	1
Colorectal Cancer	30-69	1.80 (1.46 - 2.22)	1.27 (0.86 - 1.87)	1.07 (0.95 - 1.20)	1
Females and Males	70-79	1.59 (1.28 - 1.98)	1.21 (0.80 - 1.83)	1.07 (0.92 - 1.24)	1
	80+	1.39 (1.11 - 1.74)	1.16 (0.70 - 1.92)	1.07 (0.84 - 1.36)	1
Diabetes	30-69	1.76 (1.44 - 2.16)	1.50 (0.90 - 2.50)	1.21 (0.95 - 1.54)	1
Females and Males	70-79	1.60 (1.28 - 1.99)	1.43 (0.79 - 2.58)	1.21 (0.88 - 1.66)	1
	80+	1.45 (1.10 - 1.91)	1.34 (0.63 - 2.87)	1.21 (0.74 - 1.98)	1
Ischaemic Heart Disease	30-69	1.97 (1.57 - 2.48)	1.66 (1.14 - 2.42)	1.15 (1.04 - 1.28)	1
Females and Males	70-79	1.73 (1.36 - 2.20)	1.51 (1.00 - 2.28)	1.15 (1.00 - 1.32)	1
	80+	1.50 (1.15 - 1.96)	1.38 (0.86 - 2.21)	1.15 (0.94 - 1.41)	1
Ischemic Stroke	30-69	1.72 (1.09 - 2.71)	1.23 (0.41 - 3.67)	1.12 (0.62 - 2.03)	1
Females and Males	70-79	1.55 (0.96 - 2.49)	1.21 (0.36 - 4.08)	1.12 (0.54 - 2.32)	1
	80+	1.39 (0.80 - 2.42)	1.18 (0.23 - 6.06)	1.12 (0.36 - 3.53)	1

Categories of physical activity were defined using responses to questions regarding physical activity during the past 30 days.
Source: Danaei et al., 2009 [41]

3.2 Air pollution ($PM_{2.5}$)

The GBD 2017 relative risk measures for the association between air pollution and previously modelled health outcomes (ischemic stroke, IHD, tracheal, bronchus and lung cancer and COPD) were considered the most up to date evidence for use in the NSW model. As mentioned earlier in section 1.4, additional health outcomes (lower respiratory tract infection, type 2 diabetes, intracerebral haemorrhage and subarachnoid haemorrhage) associated with air pollution considered for inclusion in the updated model, were all included in the latest update of the GBD study [16]. We applied the relative risks used in the GBD study to the *NSW Active Transport Health Model* (Table 6).

Table 6 Relative risks used by age for each outcome for the $PM_{2.5}$ (10 µg/m³), morbidity and mortality for both males and females

Outcome	Age	Relative risk Mean (95% confidence interval)
Lower respiratory infections	All ages	1.108 (1.046-1.219)
Tracheal, bronchus, and lung cancer	All ages	1.077 (1.041-1.126)
Ischaemic heart disease	25-29 years	1.238 (1.105-1.425)
Ischaemic heart disease	30-34 years	1.229 (1.101-1.402)
Ischaemic heart disease	35-39 years	1.215 (1.102-1.385)
Ischaemic heart disease	40-44 years	1.192 (1.083-1.339)
Ischaemic heart disease	45-49 years	1.183 (1.086-1.325)
Ischaemic heart disease	50-54 years	1.168 (1.077-1.293)
Ischaemic heart disease	55-59 years	1.155 (1.078-1.268)
Ischaemic heart disease	60-64 years	1.141 (1.068-1.241)
Ischaemic heart disease	65-69 years	1.125 (1.062-1.217)
Ischaemic heart disease	70-74 years	1.112 (1.053-1.190)
Ischaemic heart disease	75-79 years	1.099 (1.049-1.162)
Ischaemic heart disease	80-84 years	1.084 (1.041-1.139)
Ischaemic heart disease	85-89 years	1.070 (1.036-1.113)
Ischaemic heart disease	90-94 years	1.056 (1.030-1.095)
Ischaemic heart disease	95+ years	1.042 (1.021-1.069)
Stroke[a]	25-29 years	1.167 (1.036-1.347)
Stroke	30-34 years	1.155 (1.032-1.333)
Stroke	35-39 years	1.146 (1.024-1.303)
Stroke	40-44 years	1.133 (1.027-1.297)
Stroke	45-49 years	1.122 (1.025-1.261)
Stroke	50-54 years	1.113 (1.023-1.248)
Stroke	55-59 years	1.104 (1.024- 1.228)
Stroke	60-64 years	1.093 (1.021-1.196)
Stroke	65-69 years	1.083 (1.019-1.174)
Stroke	70-74 years	1.075 (1.017-1.157)
Stroke	75-79 years	1.064 (1.015-1.133)
Stroke	80-84 years	1.054 (1.015-1.107)
Stroke	85-89 years	1.047 (1.012-1.094)
Stroke	90-94 years	1.037 (1.010-1.073)
Stroke	95+ years	1.028 (1.008-1.057)
Chronic obstructive pulmonary disease	All ages	1.170 (1.070-1.296)
Diabetes mellitus type 2	All ages	1.282 (1.116-1.466)

Source: GBD 2017 [16]. LCI: lower boundary of 95% confidence interval; HCI: upper boundary of 95% confidence interval.
[a]Stroke includes ischaemic stroke, intracerebral haemorrhage, and subarachnoid haemorrhage

4. Assessment of the evidence for causal relationships of physical activity with the additional health outcomes considered for inclusion on the model

4.1 Rationale

One of the key issues that emerged from the stakeholder consultation done in Phase 1 was that the updated model should consider all relevant health outcomes that have sufficiently strong epidemiological evidence of an association with active transport. Towards this goal, diseases for which the association was found strong enough in the latest update of the GBD study [16] have been included in the *NSW Active Transport Health Model*. Informed by the evidence gathered in Phase 1 of the project, additional health outcomes that are not currently included in the latest GBD study were suggested for inclusion in the model. These outcomes include musculoskeletal diseases (LBP, OA) and mental health conditions (depressive disorders [shortened to 'depression' in the systematic review reporting and model], anxiety). In addition, the evidence led to a recommendation for the inclusion of all-cause mortality in the updated model to cover the whole spectrum of mortality outcomes.

Musculoskeletal diseases and mental health problems are responsible for a large burden of ill health in Australia and even a modest relationship with PA could have a significant benefit to health from interventions that improve levels of PA.

In the GBD 2017 study, an analysis of the health problems that caused the most disability in Australia ranked LBP as top on the list. Over the decade from 2007 to 2017, the burden in 'years lived with disability' (YLD) increased by 18.2%. Depressive disorders ranked third (+7.9%), anxiety disorders came in 5th (+13.7%), and other musculoskeletal diseases being ranked 7th (+8.7%) (Figure 5) [17].

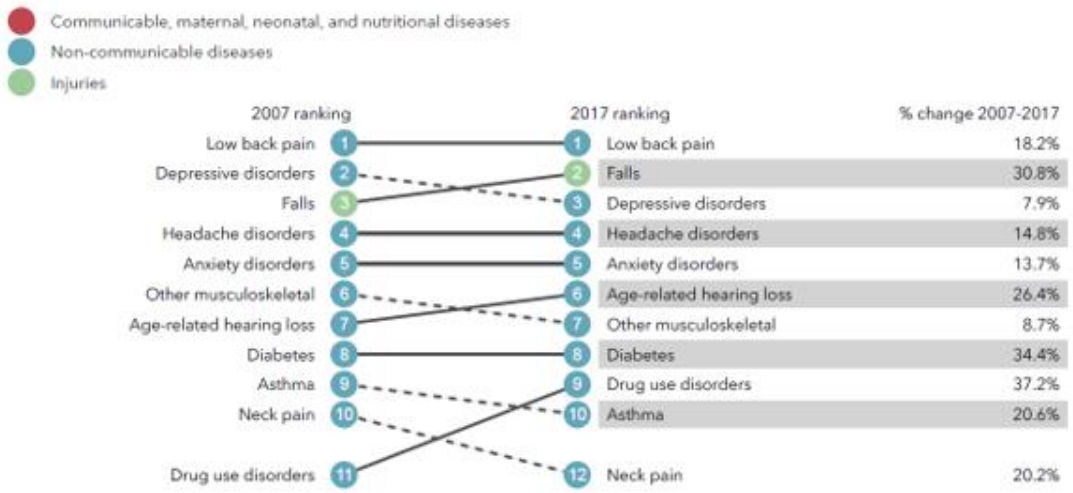


Figure 5 Top 10 causes of years lived with disability (YLDs) in 2017 and percentage change, 2007-2017, all ages, Australia
Source: Institute for Health Metrics and Evaluation. GBD Study 2017 **[17]**

In the 2011 Australian burden of disease study, a ranking of reduction of healthy life expectancy by disease group, musculoskeletal diseases and mental health recorded the largest reduction in healthy life expectancy for NSW [44]. Musculoskeletal disease led to 1.8 (males) and 2.3 (females) year reductions in healthy life expectancy. Mental health led to the loss of 1.8 (males) and 1.7 (females) healthy life years (Figure 6).

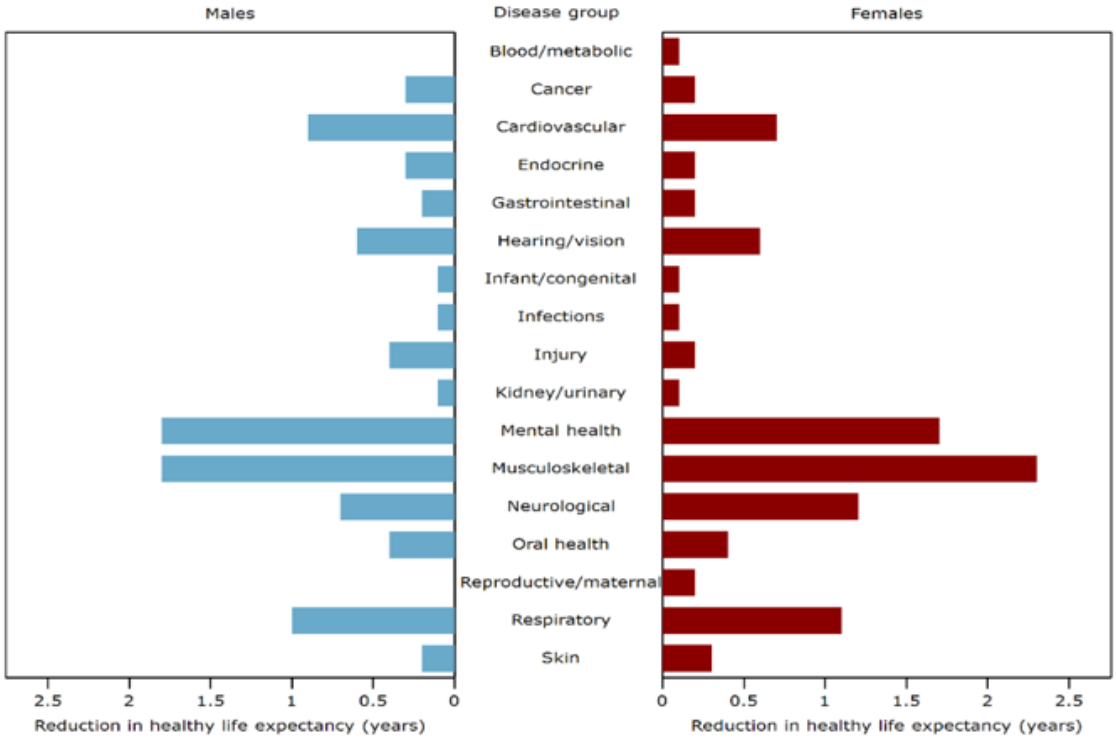


Figure 6 Reduction in healthy life expectancy (at birth) by disease group, NSW 2010-2012
Source: Centre for Epidemiology and Evidence. HealthStats NSW. [44]

Similarly, the population burden in NSW by disease group and persons in all ages category was highest for mental health (23%), followed closely by musculoskeletal disease (22.4%) [44]. This significant burden further supports the consideration of inclusion of these health outcomes in the NSW Active Transport Health model. Particularly, we include depressive disorders, anxiety, OA and LBP in the model.

To investigate whether the evidence for depressive disorders, anxiety, OA and LBP being causally related to physical activity is sufficiently strong to warrant including in the NSW Active Transport Health model, we carried out three systematic reviews to establish the epidemiological evidence on the strength of the association between exposure to PA, and assessed the likelihood of these associations being causal.

4.2 Considerations of strength of association and evidence for causal relationships

Through qualitative analysis of included studies, we established the magnitude of the associations between exposure to PA, and depression, anxiety, OA, LBP, and ACM. For anxiety and depression, the combined measures of associations are presented in Appendix K. After establishing associations, we then moved on to assess the likelihood of these associations being causal. To appraise the quality of the evidence presented in a review, we applied a modified version of GRADE (Grading of Recommendations, Assessment, Development, and Evaluation) guidelines (Appendix H) [45]. Additionally, we used the Bradford Hill criteria to assess the reviews against a causal criteria (Appendix H) [45, 46]. Thereafter, we graded the evidence to support a judgement of a relationship. This grading process was guided by the World Cancer Research Fund grading system [23]. We adopted grades of convincing, probable, possible and insufficient evidence (Appendix H) similar grading system was applied in the GBD 2017 study [22].

In this section, only a summary of the appraisal of quality of evidence and assessment for causal relationship is presented. A detailed report of the findings is recorded in respective manuscript drafts (Appendix K, Appendix L, Appendix M).

4.2.1 Physical activity and Depression

Strength of association

The systematic reviews by Mammen and Faulkner [47] and Schuch and colleagues[48] provide evidence of an association between PA and depression. Mammen and Faulkner [47] perfomed a systematic review of prospective cohort studies. They found a significant, inverse relationship between baseline PA and depression in later years in 25 of the 30 included studies, after adjusting for potential other explanatory variables. This suggests that physical activity can prevent the onset of depression. A similar finding was seen in a meta-analysis of prospective cohort studies done by Schuch and colleagues[48]. The authors found that compared with people with low levels of PA, those with higher levels of PA have lower risks of developing depression.

A summary of the measures of disease association found in our review study is presented in Appendix K. For the association between PA and depression, measures of association for use in our model were taken from the Schuch and colleagues[48] study (adjusted relative risk=0.83, 95% CI=0.76, 0.90). For modelling purposes, we assumed that the *lowest* category in their paper refers to the *inactive* category in our model, and *highest* category refers to the *highly active* category in our model. Scaling was done to interpolate values between ‘highest’ and ‘lowest’ in the studies, using the study by Ekelund and colleagues [30] which pooled the results of accelerometry-measured PA (see section 4.2.5). Figure 7 shows the resulting risk curve which depicts the relationship between PA and depression, when the PA levels shift a small distance to the right, it lowers the levels of risk.

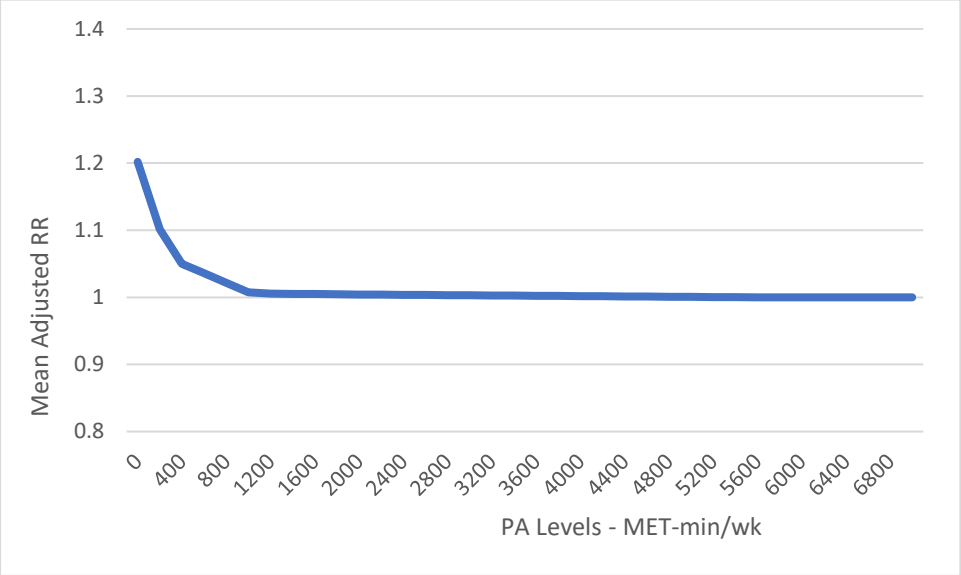


Figure 7 Relative risk of incident depression by level of physical activity
Source of relative risk values: Schuch et al., 2018 [48]

Evidence for causal relationship

Table 7 presents a summary of the assessment of the evidence for a causal relationship. A detailed report on appraisal for causal relationship is presented in Appendix K.

Table 7 Assessing the evidence against causal criteria: Depression

Criterion	Findings
1.Temporality	Pooled results from prospective studies show that lower baseline levels of PA are associated with depression at follow up [47, 48].
2. Strength of association	From our classification of measures of effect in our guidelines (Appendix H), the relative effects reported from the studies were considered weak (Appendix K). Weakness of an association makes the risk of alternative explanations greater but does not preclude causality.
3. Consistency	Consistent findings were observed across the two review studies included in our review. In their discussions, the authors of the two review studies also indicated that consistency had been witnessed in the various primary studies that had been carried out in different populations.
4. Dose-response relationship	A dose relationship was reported in various primary studies included in the review by Mammen and Faulkner [47]. A dose-response relationship can add weight to an evaluation of causation [45, 46]
5. Biological plausibility	There are various possible biological mechanisms through which insufficient PA could cause depression. We concluded that the available evidence meets the biological plausibility criteria.
6. Specificity	This criterion is not met – physical inactivity does not invariably lead to depression, and depression is not the only health condition associated with inactivity. However, this criterion was thought of in relationship to infectious agents, and seldom applies.
7.Coherence	We considered that the interpretation of the association between PA and depression does not conflict with what is known of the natural history and biology of depression.

Assessment of grade of evidence Convincing / Probable / Possible / Insufficient	When assessed against Bradford Hill's criteria for causality [45, 46], we graded the evidence as strong enough to support a judgement of a probable causal relationship, with higher levels of physical activity probably leading to a lower risk of depression. This supports the inclusion of depression in the *NSW Active Transport Health Model*.

4.2.2 Physical activity and Anxiety

Strength of association

In their recent systematic review and meta-analysis, McDowell and colleagues [49] synthesised population-based evidence of a prospective association between PA and incident anxiety disorders. The authors found that with higher PA exposure, the odds were significantly lower for self-reported anxiety symptoms, diagnosed anxiety disorder, and generalized anxiety disorder. The second study included in our qualitative synthesis was by Schuch and colleagues [50]. These authors performed a meta-analysis of prospective cohort studies to examine the relationship between PA and incident anxiety. They found that higher self-reported PA levels were associated with lower rates of incident anxiety when compared with lower PA levels. A summary of the measures of disease association found in our review study is presented in Appendix K

For the association between PA and anxiety, we used the adjusted odds ratio from the strongest available meta-analysis by Schuch and colleagues [50] (AOR = 0.74, 95% CI = 0.62, 0.88). As this paper presented results for the comparison between highest and lowest exposure categories, the assumption was made to consider these as equivalent to the *inactive* category in the model and the *highly active* category, respectively. For the intermediate categories, scaling was done to interpolate values between 'highest' and 'lowest' in the studies, using the study by Ekelund and colleagues [30] which pooled the results of accelerometry-measured PA on mortality (see section 4.2.5). Figure 8 shows the resulting risk curve.

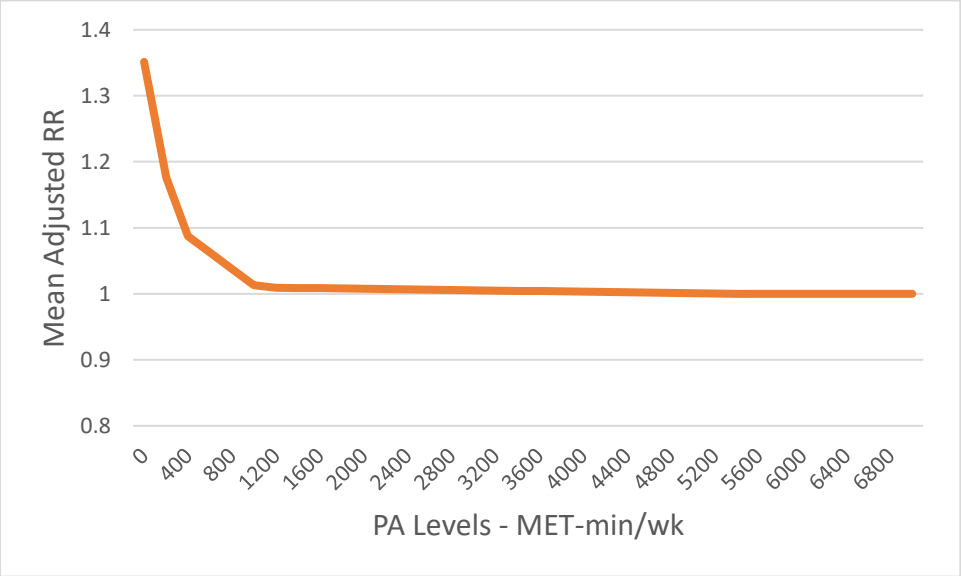


Figure 8 Relative risk of incident anxiety by level of physical activity
Source of the measures of association: Schuch et al., 2019 [50]

Evidence for causal relationship

The main findings of our evidence review are summarised in Table 8. A detailed report on appraisal for causal relationship is presented in Appendix K.

Table 8 Assessing the evidence against causal criteria: Anxiety

Criteria	Description
1.Temporality	Both reviews by Schuch and colleagues [50] and McDowell and colleagues the [49] included only prospective cohort studies with a follow up period of more than 1 year. These studies assess prospective evidence where exposure (PA) preceded the effect (anxiety) in time. This evidence meets the temporality criterion: exposure preceded outcome in the included cohort studies.
2. Strength of association	We used measures proposed by Webb and colleagues [45] to guide our classification of the strength of association between PA and anxiety. OR values >0.67 were classified as weak associations, and thus the association (OR 0.74) qualifies as weak. Though the strength of association facilitates assessment for possible causal relationship, a strong association is neither necessary nor sufficient for causality, and weakness is neither necessary nor sufficient for concluding absence of causality.
3. Consistency	In the study by McDowell and colleagues [49] study, all crude and adjusted associations included in the current meta-analyses indicated inverse associations between physical activity and subsequent anxiety. Schuch and colleagues [50] also provide additional evidence of the protective effects of self-reported PA on anxiety development referencing previous cross-sectional studies. We found that there is evidence for repeated observation of an association between PA and anxiety from other studies in different populations under different circumstances. These findings support the criterion of consistency.
4. Dose-response relationship	In their study, Schuch and colleagues [50] did not investigate a dose response relationship between physical activity exposure and anxiety. In the study by McDowell and colleagues [49], a total of 11 included studies assessed for a dose response relationship between PA and various anxiety outcomes. All the 11 studies reported lower odds of anxiety outcomes for increased amounts of PA. In all, there is modest evidence of a dose-response relationship.
5.Biological plausibility	Though the mechanisms are largely unclear, both studies presented evidence of potential biological processes that may underlie the protective effect of PA on incident anxiety.
6. Specificity	As with depression, this criterion is not met but also of questionable relevance.
7.Coherence	The interpretation for the association of PA and anxiety does not conflict with what is known of the natural history and biology of anxiety.
Assessment of grade of evidence Convincing / Probable / Possible / Insufficient	When assessed against Bradford Hill's criteria for causality [45, 46], we considered that the evidence supports a judgement of a *probable* causal relationship. This supports the inclusion of anxiety in the *NSW Active Transport Health Model*.

4.2.3 Physical activity and Low back pain

Strength of association

Two systematic review studies [51-53] reported evidence suggestive of an association between PA and LBP. Nevertheless, there were some mixed results across various variables investigated in the two studies. In a meta-analysis of seven cohort studies, Alzahrani and colleagues [51] found that compared to low level total PA, medium level total PA was significantly associated with a decreased risk of developing LBP (RR = 0.90, 95% CI 0.85 to 0.96). However, in a meta-analysis of nine cohort studies, compared to low level total PA, high level total PA was not associated with LBP (pooled risk ratio 1.00, 95% CI 0.92 to 1.08). Shiri and Falah-Hassani conducted a systematic review and meta-analysis of 36 prospective cohort studies to assess the effect of leisure time physical activity (LTPA) on non-specific

LBP [52]. Their results indicated that LTPA was neither associated with presence of LBP in the past month nor associated with presence of LBP in the past 6–12 months. However, they found that moderately and highly active individuals had a reduced risk for frequent or chronic LBP when compared against individuals with no regular PA (Table 9). A summary of the measures of disease association found in our review study is presented in Appendix I.

The adjusted risk ratios describing the association between PA and frequent or LBP in the Shiri and Falah-Hassani [52] study were selected for use in the model (Table 9). The resulting risk curve is presented in Figure 9.

Table 9 Measures of association used in the model: Low back pain

Physical activity		Adjusted risk ratios (95% CI)		
Category	MET-min/wk	Mean	LCI	HCI
None	-	1	1	1
Moderate	1025	0.86	0.79	0.94
High	3100	0.84	0.75	0.93

Measures of association between physical activity and frequent or chronic low back pain

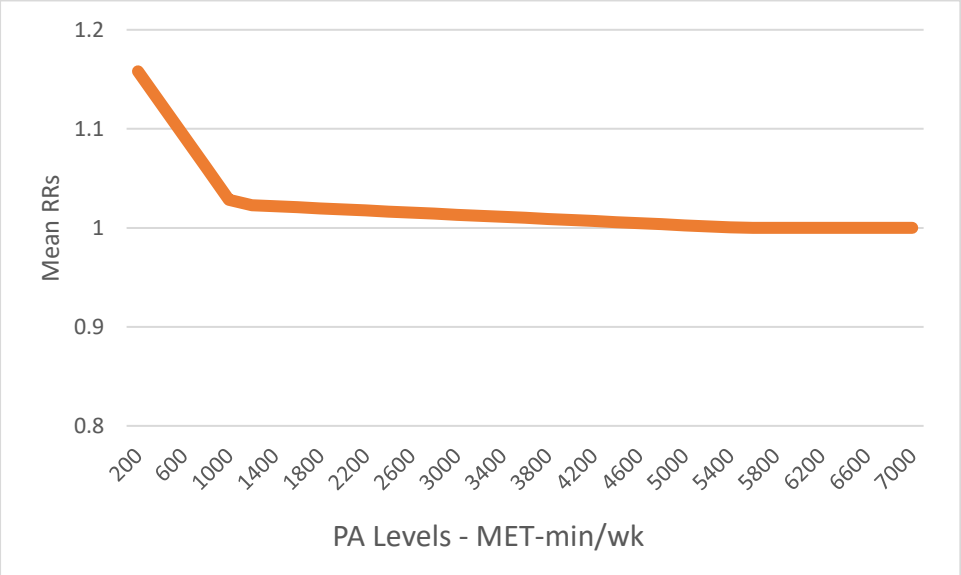


Figure 9 Relative risk of incident low back pain by level of physical activity
Source of relative risk values: Shiri & Falah-Hassani, 2017 [52]

Evidence for causal relationship

Two studies investigating the association between physical activity and low back pain met the inclusion criteria for our qualitative synthesis [51, 52] . The main findings of our evidence review for causal relationship are summarised in Table 10. The detailed report is presented in Appendix L.

Table 10 Assessing the evidence against causal criteria: Low back pain

Criteria	Description
1.Temporality	From the two systematic review and meta-analyses studies included in our study [51, 52], we appraised evidence from prospective cohort studies. Inclusion of prospective studies limited the possibility of reverse causation bias allowing for the examination of the temporal sequence between baseline levels of PA and LBP. We considered the evidence from these studies to have met the temporality criteria where the exposure (PA) was seen to precede the effect (LBP).
2. Strength of association	The results for the associations specific associations found between PA and LBP are reported in detail in Appendix I. The findings provide evidence that suggests that there is an inverse association between PA and LBP. For instance in the study by Shiri & Falah-Hassani [52], the risk of frequent/chronic LBP was 14% lower (RR=0.86, CI 0.79 to 0.94, I2=0%, n=33 032) in moderately active individuals and 16% lower (RR=0.84, CI 0.75 to 0.93, I2=0%, n=33 032) highly active individuals in comparison with individuals without regular physical activity. Using an adaptation of measures by Webb and colleagues [45] to guide our classification of the strength of association, we considered the reported measures of strength of association weak. Weakness of an association makes the risk of alternative explanations greater but does not preclude causality.
3. Consistency	Our findings provide cautious support for the criterion of consistency. Alzahrani and colleagues [51] concluded that their results provided evidence suggesting that there is an inverse association between physical activity and LBP. Compared to low level total physical activity, medium level total physical activity was significantly associated with a decreased risk of developing low back pain (fully adjusted risk ratios, 0.90, 95% CI 0.85 to 0.96). Shiri and Falah-Hassani [52] considered their findings to suggest that moderate to high level of physical activity during leisure time protects against frequent or chronic LBP.
4. Dose-response relationship	In their systematic review and meta-analysis, Alzahrani and colleagues [51] found no evidence of dose-response relationship. Shiri and Falah-Hassani [52] did not investigate a dose response relationship.
5.Biological plausibility	There are several proposed biological mechanisms by which PA is associated with LBP. Though strenuous PA has been considered to increase the risk of LBP, the kind of PA involved in active transport is unlikely to have the same impact. Instead, this type of PA may lead to increased muscle strength and flexibility hence protecting the spine from injuries.
6. Specificity	Not supported but not highly applicable.
7.Coherence	The interpretation for the association of PA and LBP does not conflict with what is known of the natural history and biology of LBP.
Assessment of grade of evidence Convincing / Probable / Possible / Insufficient	From the consistent findings in the two review studies included in our qualitative synthesis [51, 52], there was sufficient evidence indicative of an association between PA and LBP. When assessed against Bradford Hill's criteria for causality we graded our findings as possible (suggestive) evidence for a causal relationship. This supports the inclusion of LBP in the *NSW Active Transport Health Model* but it should be reserved for sensitivity analyses and not be included in the main analysis.

4.2.4 Physical activity and Osteoarthritis

Strength of association

The two studies identified for inclusion in our review for the outcome OA [54, 55] did not yield combined measures of association that we could use for our study. Hart and colleagues [54] included ten cohort studies in a review that examined the relationship between physical activity and the development or progression of OA in later life. These authors provided a narrative report of findings. We reviewed each of the ten included studies to see if they would provide us with evidence on combined measures of association and hence facilitate our assessment for possible causal relationship between PA and osteoarthritis. However, we did not consider any of the studies suitable for inclusion in our study. Some of the reasons for this were; the study populations were not representative of the general population; two studies focused on subjects who were identifiable exclusively as athletes while one study focused on patients, one study provided no measures of association for the relationship between PA and OA, a total of six studies investigated very high levels of PA only. From their findings, Hart and colleagues [54] considered high levels of PA a risk factor for OA. The second study was a systematic review by Richmond and colleagues [55]. The authors sought to establish whether joint injury, sport activity, physical activity, obesity, or occupational activities were predictors for OA of the knee, hip, and ankle. Five articles met the inclusion criteria for the physical activity aspect (3 cohort studies and 2 case-control studies). We further reviewed the 5 individual articles included in the review by Richmond and colleagues [55] and found that we could not include these articles in our review. This was because study populations were not representative of the general population (n=2), one study exclusively looked at PA at work, and two were case control studies, one of which had a definition of PA exposure that was different from that used in our study. The authors found mixed results on the relationship between physical activity and OA. However, they concluded that the evidence was not strong enough to support an association between increased exposure to sport and/ or physical activity and increased risk of knee or hip OA [55].

The varying results found in these studies could be due to different types of PA having different effects on joint health. Repeated intense stress on the joints, or acute high stress, which can occur in manual labour or intense sport training, may damage cartilage, ligaments and other joint structures, which may lead to OA. In contrast, the kind of PA involved in active transport is unlikely to have the same impact, but instead stabilise the joints by strengthening muscles. This may explain the findings in the study by Alzahrani and colleagues [51] that only moderate PA is associated with a lower risk of LBP.

We carried out an additional review of literature using a modified protocol to identify cohort studies investigating the association between walking, cycling or 'active transport' and OA. The strongest study we identified was a prospective cohort study by White and colleagues [56]. This study examined the association of step-defined daily walking with incident functional limitation two years later in people with or at risk of knee OA (Appendix L). The authors found that people who walked more were less likely to have developed functional limitations two years later. The measures of effect presented in this study (Table 11) were selected for use in the model because of its strong design and exposure measurement, but recommend it be used only in sensitivity analyses, not the main analysis. We assumed that the risk of functional limitation equals the risk of developing OA. OA is usually diagnosed based on both functional limitations and imaging (X-rays or MRI). Figure 10 presents the risk curve.

Table 11 Measures of association used in the model: Osteoarthritis

Walking & Incident functional limitation 2 years later (n=1788)					
Steps/ day	Baseline [Mean]		Adjusted RR		
	Gait Speed (m/s)	MET-min/wk [a]	Mean	LCI	HCI
< 5000	1.19	428	1	1	1
5000 – 7499	1.26	1,012	0.5	0.33	0.76
7500	1.32	1,545	0.31	0.18	0.56

[a] Calculated value for use in the model

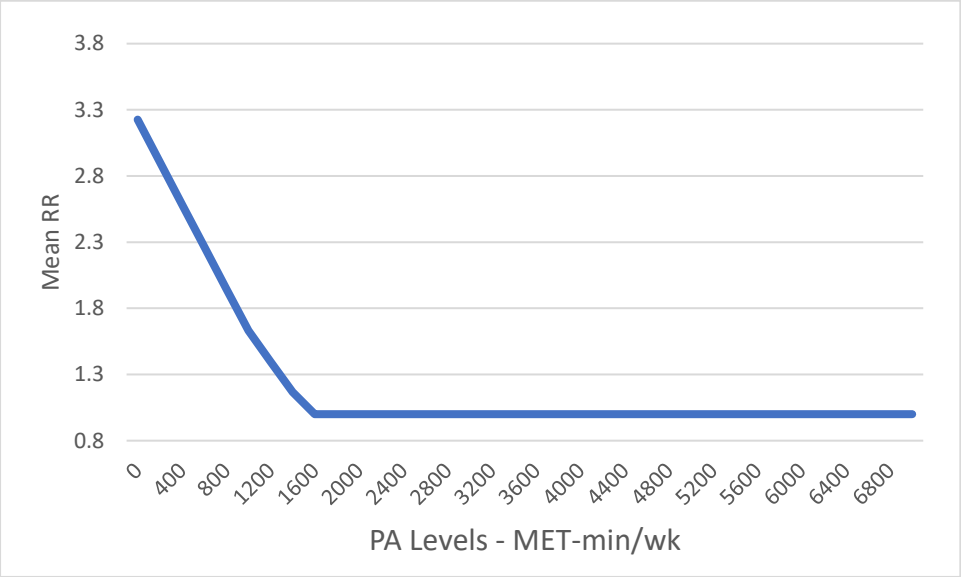


Figure 10 Relative risk of incident osteoarthritis by level of physical activity
Source of relative risk values: White et al., 2014 [56]

For our model, we used the reported adjusted risk ratios of incident functional limitation at the two-year follow-up as measured by gait speed during the 20-meter walk among the study participants. We used the reported steps/day to estimate the steps per week category. We considered that each step was 0.76 meters [57]. Using the gait speed (meters walked per second) reported in the findings, we converted the meters walked per week into hours spent walking per week. Applying a MET value of 2.30 for walking compared to sitting [29], we then calculated the MET-minutes/week for each category of PA (steps/day).

Evidence for causal relationship

The main findings of our evidence review for causal relationship are summarised in Table 12. The detailed report is presented in Appendix L.

Table 12 Assessing the evidence against causal criteria: Osteoarthritis

Criteria	Description
1.Temporality	In the large cohort study by White and colleagues [56], data was pooled data from 6 prospective cohort with a median follow up time of 14.2 years. Only study participants without functional limitation at baseline were included. Inclusion of prospective studies limited the possibility of reverse causation bias allowing for the examination of the temporal sequence between PA and functional limitation. Since OA is usually diagnosed based on both functional limitations and imaging (X-rays or MRI), functional limitation was taken to represent OA in our study. We considered the evidence to have met the temporality criteria where the exposure (PA) was seen to precede the effect (OA).
2. Strength of association	The findings by White and colleagues [56] indicated that a greater number of steps/days, measured either by self-report or performance based, was increasingly protective against the development of functional limitation. Using an adaptation of measures by Webb and colleagues [45] to guide our classification of the strength of association, the reported measures of strength of association ranged from modest to moderately strong. Weakness of an association makes the risk of alternative explanations greater but does not preclude causality.
3. Consistency	The findings were consistent with previous studies discussed by the authors. The authors also highlight that the PA thresholds for functional limitation found in the study were consistent with the findings for other clinical outcomes such as adverse cardiometabolic health indicators.
4. Dose-response relationship	The findings give evidence of a dose response relationship between PA and functional limitation
5.Biological plausibility	In our study, we assumed that the risk of functional limitation equals the risk of developing OA. Repeated intense stress on the joints, or acute high stress, which can occur in manual labour or intense sport training, may damage cartilage, ligaments and other joint structures, which may lead to OA. In contrast, the kind of PA involved in active transport is unlikely to have the same impact, but instead stabilise the joints by strengthening muscles.
6. Specificity	Not supported, not readily applicable.
7.Coherence	The interpretation for the association of PA and OA does not conflict with what is known of the natural history and biology of OA.
Assessment of grade of evidence Convincing / Probable / Possible / Insufficient	We graded our findings as possible (suggestive) evidence for a causal relationship. This supports the inclusion of OA in the *NSW Active Transport Health Model*. However, the earlier assumption made on the outcome, functional limitation being equal to OA, is a limitation to our study. We therefore propose to include the OA only in the sensitivity analysis for the model.

4.2.5 Physical activity and All-cause mortality (ACM)

Strength of association

The eight systematic review and meta analyses studies included in our qualitative synthesis provided us with sufficient evidence supporting the association between PA and ACM [30, 58-64]. All eight studies included only prospective cohort studies in their analyses. Where the authors presented pooled effect sizes, we summarize the measures of disease association in our findings (Appendix J). In this section, we present a narrative synthesis of our findings on the strength of association between PA and ACM.

In a systematic review and harmonized meta-analysis study, Ekelund and colleagues [30] investigated the dose-response associations between accelerometry measured PA and ACM. They found that higher levels of total PA were associated with substantially reduced risk for premature mortality (Table 13). Also, a non-linear dose-response pattern was evident in their findings. Hupin and colleagues [59], carried out a systematic review and meta-analysis to determine whether moderate-to-vigorous-intensity physical activity (MVPA) lower than the current PA recommendations was effective in reducing mortality. They found that when compared with participants in the inactive group, participants with a low dose of MVPA (1–499 MET-min per week), had a 22% lower mortality risk (Appendix J). MVPA beyond this threshold improved these benefits in a linear fashion. Lollgen and colleagues [61] reported lower ACM for active individuals (Appendix J). A dose response relationship was evidenced by a marked reduction in mortality with light and moderate activities, and a small additional risk reduction seen for vigorous exercise intensity. A systematic review and meta-analysis by Woodcock and colleagues [64] found that being physically active reduced the risk of ACM. These authors highlight that the largest benefit was found from moving from no activity to low levels of activity (**Error! Reference source not found.**, Appendix J). In their review, Nocon, Hiemann [62] also c oncluded that being physically active reduced the risk of ACM. The review study by Samitz and colleagues [63] quantified the relationships between ACM and different domains of PA. The authors found that higher levels of total and domain-specific PA were associated with reduced ACM (Appendix J). Two studies reported associations for walking as PA exposure [58, 60]. Of the two studies, one investigated the association between cycling and ACM [60]. These two studies reported inverse relationships between the exposures (cycling and walking) and ACM. Additionally, Kelly and colleagues [60] presented evidence of a dose-response that suggested decreasing rate of benefit at higher PA exposure. The measures of associations reported in these studies are summarised in Appendix J.

For the association between PA and ACM, three sets of relative risk measures were selected for use in our model. The first set was taken from the paper by woodcock and colleagues [64]. We used the author's central estimate of moderate PA exposure levels (Table 13). We contacted James Woodcock seeking additional data used to quantify the associations between PA and ACM in their paper. In response, the author recommended a population-based prospective cohort study by Arem and colleagues [65]. Arem and colleagues used pooled data from 6 cohorts in the National Cancer Institute Cohort Consortium (baseline 1992-2003) to quantify the dose-response association between LTPA and mortality. They also sought to define the upper limit of benefit or harm associated with increased levels of PA. The authors reported that compared with no baseline LTPA, any level of activity was associated with a significantly lower risk of mortality (Table 13). These values were proposed for use in our model.

Also, we selected to include measures from the study by Ekelund and colleagues [30] that examined the dose-response associations between accelerometer assessed total PA and ACM (Table 13). The risk curves are shown in Figure 11.

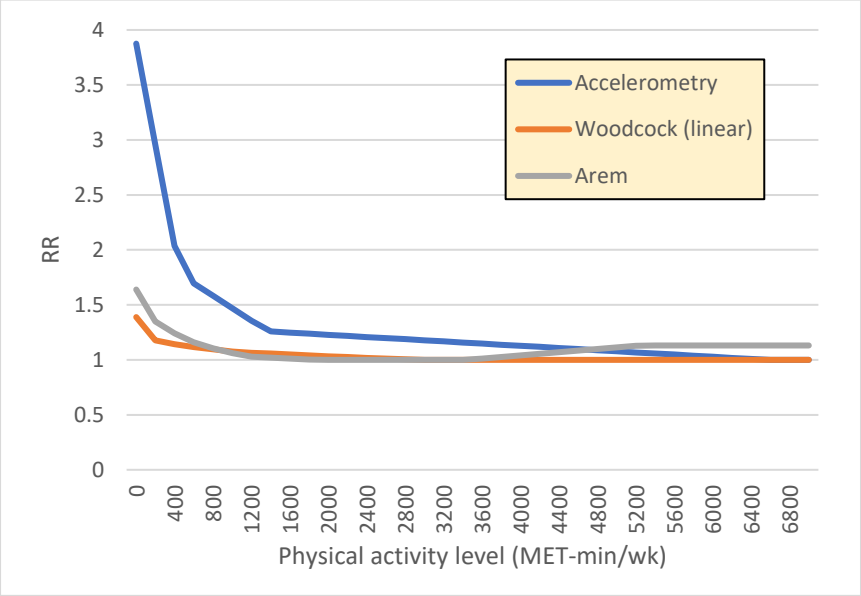


Figure 11 Relative risk of mortality by level of physical activity
Source of the measures of association: Woodcock et al., 2011 [64], Arem et al., 2015 [65] & Ekelund et al., 2019 [30]

Table 13 Measures of association used in the model: All-cause mortality

Woodcock et al., 2011 [64]				
PA		RR, with low and high boundaries of 95% confidence intervals (CIs)		
Hr/wk	MET-min/wk	Mean	LCI	HCI
0	0	1	1	1
1	210	0.84	0.81	0.88
2.5	525	0.81	0.76	0.85
5	1050	0.77	0.73	0.82
7	1470	0.76	0.71	0.81
10	2100	0.74	0.68	0.79
14	2940	0.72	0.66	0.78

Arem et al., 2015 [65]				
		Hazard ratios (HRs) and 95% confidence intervals (CIs) for leisure time moderate- to vigorous-intensity physical activity (PA) and mortality [a]		
PA MET h/wk	PA MET Min/wk	HR Mean	LCI	HCI
0	0	1	1	1
0-<7.5	0 - <450	0.8	0.78	0.82
7.5-<15.0	450 - <900	0.69	0.67	0.7
15.0-<22.5	900 -<1350	0.63	0.62	0.65
22.5-<40.0	1350-<2400	0.61	0.59	0.62
40.0-<75.0	2400 -<4500	0.61	0.58	0.64
75.0+	4500+	0.69	0.59	0.78

Ekelund et al., 2019 [30]				
PA		Hazard ratios (95% CI) for all-cause mortality		
Quarter	Cpm	Mean	LCI	HCI
First quarter (least active)		1	1	1
Second quarter	168	0.48	0.43	0.54
Third quarter	256	0.34	0.26	0.45
Fourth quarter (most active)	335	0.27	0.23	0.32

[a] Multivariable-adjusted models are adjusted for gender, smoking (never, former, current, missing), alcohol (none, <15 grams/day, 15-30 grams/day, 30+ grams/day), education (dropout, high school, post high school education, some college, college graduate, post-college, missing), marital status (married, divorced, widowed, single, missing), history of cancer, history of heart disease, and where indicated, body mass index (<18.5, 18.5-25, 25-<30, 30-<35, 35+ kg/m2). cpm = counts per minute.

Evidence for causal relationships

In this section, we present the appraisal of evidence and assessment of causal relationship between physical activity and all-cause mortality. Eight systematic review studies [30, 58-64] investigating the association between physical activity and all-cause mortality met the inclusion criteria for our qualitative synthesis. Additionally, as mentioned earlier in our report, a population-based prospective cohort study by Arem and colleagues [65] was included in the appraisal of the strength of association and possible causal relationship between PA and ACM. A summary of this appraisal of evidence for a causal relationship is summarised in Table 14. A detailed appraisal is presented in Appendix M.

Table 14 Assessing the evidence against causal criteria: All-cause mortality

Criteria	Description
1.Temporality	All the eight review studies [30, 58-64]and one large cohort study [65] considered in our qualitative synthesis included only prospective cohort studies with long follow up periods (See Appendix M). The evidence met the temporality criterion.
2. Strength of association	Using an adaptation of measures by Webb and colleagues [45] to guide our classification of the strength of association, the reported measures of strength of association ranged from weak association to moderately strong. When compared with other studies, the observed magnitude of risk reduction in the study by Ekelund and colleagues [30] is more than twice as large as previous studies that assessed PA by self-report. The strength of association facilitates assessment for a possible causal relationship. Weakness of an association makes the risk of alternative explanations greater but does not preclude causality.
3. Consistency	The consistency of results across all the included studies supports the overall conclusion of a beneficial effect of PA. The authors also discuss various comparable findings of risk reductions reported in previous studies.
4. Dose-response relationship	Authors in all the eight included systematic review studies found evidence of a dose response relationship between PA and ACM.
5.Biological plausibility	Various possible biological mechanisms that would contribute to the reduction in the risk of premature death associated with physical activity were considered. This included the favourable changes in cardiovascular risk factor profiles and improvements in endothelial function that result from PA. Reductions in cancer mortality seen with increased PA may be related to reduced fat stores, increased energy expenditure, changes in sex hormone levels, improved immune function, reductions in insulin levels and insulin-like growth factors, and reduced generation of free radicals. Further, in elderly people, it was discussed that regular physical activity reduced the risk of falls, osteoporotic fractures and disability, which in turn might reduce mortality.
6. Specificity	Not supported, not highly applicable.
7.Coherence	The interpretation for the association of PA and ACM does not conflict with what is known of the natural history and biology of the all-cause mortality outcome.
Assessment of grade of evidence **Convincing / Probable / Possible / Insufficient**	We graded our findings as convincing evidence for a causal relationship. This supports the inclusion of ACM in the *NSW Active Transport Health Model.*

5. Health outcome measures

Based on the evidence gathered in the three systematic reviews the NSW Active Transport Health Model has been extended to include mental health problems (depression, anxiety), and ACM as outcomes associated with PA (Figure 12). Musculoskeletal diseases (OA, LBP) have also been incorporated in the model but proposed for inclusion in the sensitivity analysis. Double counting of health effects is avoided by the life table structure, and by ignoring disease-specific mortality when an impact of PA on ACM is applied. This is explained further in Appendix B.

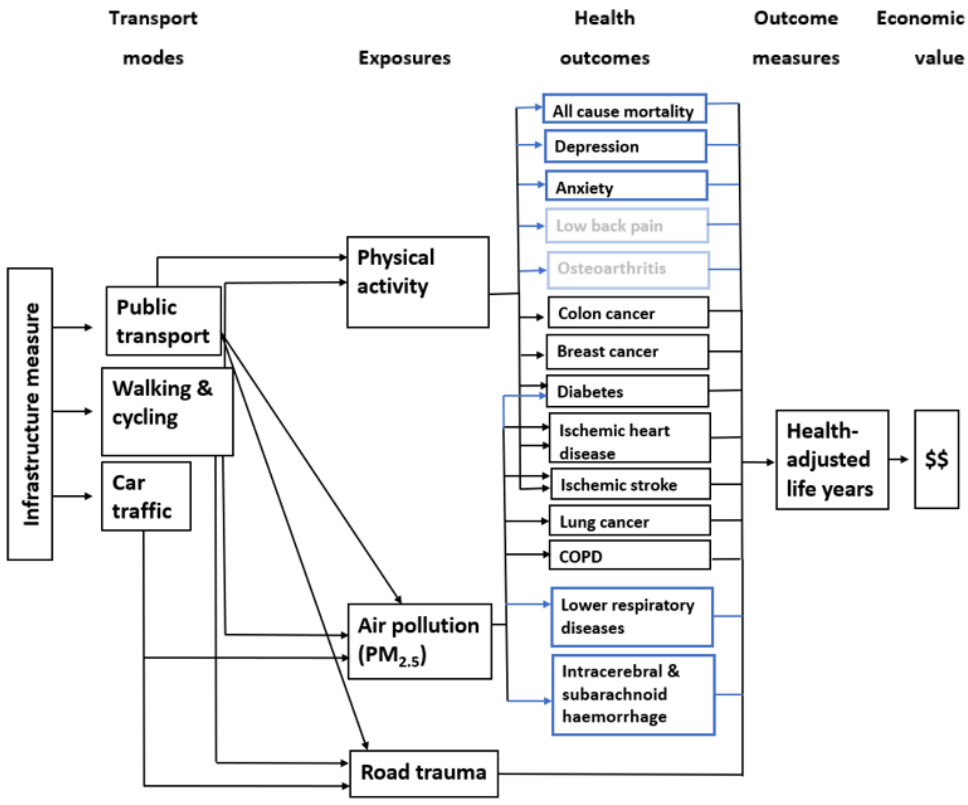


Figure 12 Schematic overview of the NSW Active Transport Health Model

The Zapata-Diomedi model measured health outcomes as life years, health adjusted life years (HALYS), prevalent cases, deaths, years lived with disability (YLDs) and health care costs. In the newly created *NSW Active Transport Health Model*, the range of outcomes has been extended greatly; the model output now includes incidence, prevalence, mortality, and health care costs over the first 25 years following the start of the intervention.

6. Valuing health-adjusted life years

The widely accepted epidemiological modelling approach we used to estimate the health effects of changes in active transport behaviours produces results that are summarised in 'health-adjusted life years' (HALYs). Estimating the economic value of these benefits requires a conversion to monetary values.

As detailed in our previous report [66], there are various methods to estimate the social value of an additional year of life in full health, or the equivalent in averted loss of quality of life, and studies using these methods produce a wide range of estimates of that value.

The NSW Active Transport Health model can accommodate any value, but for this report we produced results for the three values most commonly used in the Australian context: the Australian Transport Assessment and Planning (ATAP)/TfNSW 'Inclusive Willingness to Pay' value [1], the Office of Best Practice Regulation (OBPR)/Abelson value [67, 68], and a value commonly used in health economics literature [69]. The first two are derived from the 'value of a statistical life' (VSL). The latter is a more direct estimate of willing to pay for one year in full health ('quality-adjusted life year' or QALY).

6.1 The Inclusive Willingness to Pay approach used by TfNSW

The Inclusive Willingness to Pay approach is recommended by the Australian Government Department of Infrastructure and Regional Development and is used by Transport for NSW [1]. The method is based on NSW evidence: it uses willingness to pay estimates from a review that was prepared in 2008 for the (then) Roads and Traffic Authority of NSW by PricewaterhouseCoopers (PWC) in conjunction with the Hensher Group [67]. Hensher and colleagues used discrete choice experiments in which participants selected a travel route with multiple aspects, including the risk of road traffic injury. The willingness to pay for a reduction in the risk of death or disability from a crash was statistically derived. The 'inclusive' refers to the addition of vehicle, emergency and other crash related costs, on the argument that individuals would not have included those costs in their valuation. These additional costs are not directly health-related and are not included in the present study. This leaves willingness to pay values of $7.72 million per casualty (VSL) in 2019 terms [70]. Using the customary assumption of a loss of 40 years for every death and Treasury's recommended 7% discount rate [2], this translates to a VSLY of $578,575 in 2019 terms ($333,700 at 3%).

6.2 The OBPR value

The OBPR/Abelson value is based on a review of Australian and international empirical studies by Abelson, 2008 [71]. It is recommended by the Commonwealth Office of Best Practice Regulation and widely used in Australia (other than in the transport sector) [68]. The evidence base is broad, spanning across a range of policy areas, risk factors and health conditions, but new studies have been published since Abelson conducted his review over a decade ago and the value is rather loosely based on the evidence available at the time. It suggests a value of $4.9 million for VSL and, based on a discount rate of 3%, a value of $213,000 for VSLY in 2019-dollar terms.

6.3 Values used in health economics literature

In the initial sensitivity analysis, we also used a value of used $50,000 for a healthy life year. This is often used in health economics literature to value life years gained by health and medical interventions – and has been for about two decades, without adjustment for inflation [72]. Its origins

are rather nebulous. Five years ago, Karnon and colleagues referred to decisions by the Pharmaceutical Benefits Advisory Commission on whether or not to recommend new drugs for reimbursement and mention a value of between $45,000 and $75,000 [69]. Adjusted for inflation, the lower estimate comes close to the $50,000 used in the sensitivity analysis in this study. This value has been included in this report for completeness but is not a viable option for the NSW Active Transport Health model as it relates to health care provision, not transport infrastructure investment.

6.4 The most appropriate value for use in NSW

We used the OBPR value [68] for the reference case presented in this report, and present results for the TfNSW method in the sensitivity analysis. Several considerations deserve mention: what is included in each measure and the role of risk avoidance, the remit of the different valuation approaches, coherence across NSW government, and the most appropriate discount rate for future health impacts.

6.4.1 Valuing HALYs

The value attached to a year in full health can include various aspects, as outlined in the report on Phase 1 [66]. It can include the value of duration of life, improvement in quality of life, but related indirect benefits could also be included, such as the value of avoided health care costs, the loss of production related to loss of health or lifetime, and the intangible losses associated with death or disease that accrue to others, such as relatives or friends. It is not always clear what is included in the above values. The TfNSW value is based on subjective preferences and given by the amount of money that individuals are willing to pay for reducing the risk of their premature death while performing a certain risky activity. This means that this value includes the value of risk avoidance [73]. This is entirely appropriate when applied to road traffic deaths, but in the method proposed in this project, the value of a HALY is also applied to deaths from chronic conditions related to physical activity and air pollution, and to loss of quality of life caused by these conditions. Is it justified to apply the willingness to pay for the risk of sudden death to these conditions? ATAP recommends using the OBPR value for chronic disease-related deaths. Our method also converts a value for a sudden event (death on the road) to a stream of value over years (the value of a HALY), where the risk avoidance 'premium' is better characterised as a one-off loss. Future analyses could endeavour to separate the value of risk avoidance from the value of loss of quality and duration of life and apply this value to deaths (rather than the resulting stream of HALYs lost).

The remit of the three values differs. The value used in health economics literature is applied to interventions in health care and prevention, the TfNSW value is applied in transport, and the OBPR value is applied across government interventions. The values derive partly from different 'traditions' and to our knowledge, a comprehensive comparative review has not been made. While Hensher's studies (upon which the TfNSW method is based) would probably have been included in Abelson's review (which underpins the OBPR value) had it been available in 2008, the health economics values stand out as different. The lack of a strong evidence base for this loosely used reference value makes it difficult to assess exactly what the considerations were. Of the three values discussed here, it seems least appropriate for use across NSW Government.

Further, there are considerations of coherence. The TfNSW value is consistently used for transport business cases, and the OBPR value for analyses in other areas. This is coherent if the TfNSW value is applied to road traffic injury and death, and the difference between both values is assumed to reflect the value of risk avoidance of sudden death or disability on the road. However, in this project, one

value is applied across both road traffic injury/death and non-communicable disease, with most of the effect from the latter.

Following discussion in the project advisory group, we used the OBPR value of $213,000 per statistical life year in our reference case, mainly to remain consistent with current Treasury guidelines [68] and because most of the health gains are in the avoidance of non-communicable disease, rather than road traffic injury and fatality for which the TfNSW value was designed. Future work could use a separate value for risk avoidance to road deaths, in combination with the OBPR value for HALYs.

6.4.2 Discounting

The health gains from active travel, especially the extra life years among younger cohorts, can be decades into the future. The rate at which future health outcomes are discounted therefore exerts a powerful influence. In the *NSW Active Transport Health Model,* the discount rate is applied to three future streams of benefit (or harm): health care costs, health outcomes (HALYs), and to the conversion of the value of a statistical life to the value of a statistical life *year*.

NSW Treasury's 'Note on the Discount Rate in the Context of Health' (Supplementary File 1) explains that "there are two main classes of discount rates: the opportunity cost of capital and time preference rates. The opportunity cost of capital is the value forgone by making an investment. A time preference rate reflects the value of consumption at different points in time. The opportunity cost of capital is generally higher than time preference rates because it allows for the return to market risk and the role of taxation which do not apply to estimates of time preference rates." The opportunity cost of capital includes 'market risk' (i.e., loss of the investment). Treasury uses a 7% discount rate for the opportunity cost of capital, and 3% as the time preference rate [68].

Health care costs are financial costs, hence the opportunity cost of capital applies. These should be discounted at the 7% as per Treasury guidelines [2]. However, in the recent past health care costs have tended to rise faster than the consumer price index and one way to deal with this is to apply a lower discount rate to account for this. The present work has not done so and assumed stable health care costs in future years.

The conversion of VSL to VSLY is based on the assumption that on average, the deaths that were valued in the literature equate to 40 years of life lost. This is an assumption, not an empirically established value. A discount rate is then applied to estimate the 'net present value' of each life year. The question is what discount rate is appropriate. In our initial reading of NSW Treasury guidelines, the research team applied a 7% discount rate, which recommends a rate of 7% including for health and environmental effects [2]. For the VSL of $4.9 million, the VSLY with a 7% discount rate is $400,000. Based on Treasury NSW's advice, and after discussion in the project advisory group, the research team adopted a 3% discount rate instead, on the argument that this is the rate individuals apply to value their own health and lifetime in future years and hence the rate for private social time preference applies. This results in the VSL of $213,000 given in OBPR's Best Practice Regulation Guidance Note on the Value of statistical life [68], and that was used in the reference case in this research.

Complete agreement was not reached on the discount rate for future health benefits. Although it might seem that the public's time preference rate of 3% would apply, as in the conversion of VSL to VSLY, it was argued that the 7% discount rate for the opportunity cost of capital applies, because "future generations are generally better off when government makes investments with the highest rates of return, where the surpluses / returns can in principle be reinvested, rather than investments

with lower rates of return" [74]. Suppose that health was discounted at 3% but the investment (in active transport infrastructure) at 7%. In that case, the cost-benefit ratio is more favourable if that investment is postponed till next year, when the costs are 7% lower (seen from the present) but the benefits only 3% lower. Extending this argument further, it means that a rational decision maker postpones this investment indefinitely. In the health economics literature, this 'postponing paradox' proposed by Keeler and Cretin (which flows from the related 'consistency argument') has been used to argue against lower discount rates for health compared to costs [75].

Other authors, however, disagree. They point out that historically, willingness to pay for health has increased over time, and it has been argued that this should be incorporated in CBA [76]. Conversely, UK Treasury [77] has argued that social time preference-based discount rates include a 'wealth effect' – the expectation that we will be richer in future. This devalues future good and services – but not health. Both can be accounted for by lowering the discount rate applied to future health outcomes [78]. Further, the application of a 7% discount rate to health may not adequately reflect the expectations or preferences of those governed. This is evidenced by the accepted social discount rate of around 3%. Social-time-preference rates (as opposed to rates based on the opportunity cost of capital) are used as the basis for discounting in the UK, France and Sweden [79] . An intergenerational argument also comes into play here, to the extent that decisions about active transport infrastructure affect future generations, as is the case for children and transport to school. A normative position could be taken that governments should plan for the long-term wellbeing of the community, which suggests a lower social discount rate than individuals apply to their own future health. Some authors take the normative position that in order not to discriminate against future generations, a social discount rate of zero should be applied to health and environmental outcomes [80]. The application of a 7% discount rate by Government does the opposite: it assumes governments are (/should be) more myopic than their citizens in valuing future health outcomes.

The 7% discount rate stems from an era in which the opportunity cost of capital was much higher than it is now. As a recent report by the Grattan Institute explains, the opportunity cost of capital is related to government borrowing costs, notably by the 10-year Commonwealth bond rate [79]. When the 7% opportunity cost of capital first appeared in guidelines in 1989, this rate was close to 7%, but it has steadily decreased to around 1% in 2017. In 2018, Attema and colleagues gave the example of the long-term cost of borrowing for Canadian provinces which, adjusted for inflation, came to 1.5% [78]. In the current post-COVID era, interests paid on government bonds are below inflation rates [81], making borrowing effectively free, and the expectation is that interest rates will remain low for the foreseeable future [82]. This would suggest a lower rate for the opportunity cost of capital than the 7% currently used. This would apply to both costs and health benefits, but since the costs of investments in active transport infrastructure are mostly in earlier years than the health benefits (which stretch over decades into the future), a lower discount rate would still favour such investments.

The Pharmaceutical Benefit Advisory Committee and Medical Services Advisory Committee use a discount rate of 5% for both health outcomes and costs, without explicit justification (Supplementary File 2).

In sum, there are mixed views about the appropriate discount rates for health outcomes and the current application of the 7% discount rate for the opportunity cost of capital seems to lack a solid empirical or normative basis. In this report, in the reference case a discount rate for health outcomes of 3% is used (consistent with the social time preference used to convert VSL to VSLY). In the sensitivity analysis of this study, the effects of a uniform 7% rate and various other (combinations of) discount

rates are explored. In further work, a fundamental discussion about discount rates for health (and environmental) outcomes is recommended.

7. Model results, sensitivity analysis and the value of a kilometre walking or cycling

The *NSW Active Transport Health Model* (Figure 12) has an interface that facilitates running analyses with different settings.

Dashboard

Before running the model, enter/check the parameter settings

Values in bold red font denotes are used in the model

Bold black for making selections

Intervention			
Change can be made via mode share (population level) or change in number of trips per week for a number of persons (across ages and sex, as per current mode share - 'he who has, shall be given').			
Choose '1' for mode share, or '2' for change in number of trips for a number of persons			**2**

Transport mode	**Baseline mode share**	**Intervention n**	**Change in number of trips pppw**
Private vehicle (driver/passenger)	71.9%	71.9%	-5
Walking	14.6%	**14.6%**	**5**
Cycle	1.9%	**1.9%**	**0**
Public Transport	11.5%	**11.5%**	**0**
Number of persons involved			**10,000**

Target population			
Age	Start		**20**
	End		**100**
Impact by activity level	0 =	All categories (default)	**0**
	1 =	Inactive only	
	2 =	Low active only	
	3 =	Moderately active only	
	4 =	Highly active only	

Economic values			
Discounting	Costs	**7%**	
	Health effects	**3%**	
Value of Life			
	Value of HALY1	**$ 213,000**	OBPR guidance note, Aug 2019
	Value of a Statistical Life 2	**$ 7,713,394**	TfNSW 2018 Guidelines econ appraisal
	Value of HALY2 (3% disc conversion TfNSWVSL to VSLY)	$ 333,700	
	Value of HALY3 (7% disc conversion OBPRVSL to VSLY)	$ 367,545	

Risk factors, diseases, mortality			
Physical activity	0=off, 1=on.	**1**	
% of added travel PA NOT offset by reduction in other PA		**100%**	(100% = all travel PA is additional)
Air quality ($PM_{2.5}$) - ambient	0=off, 1=on.	**1**	
Air quality ($PM_{2.5}$) - traffic participants	0=off, 1=on.	**1**	
Road traffic injury	0=off, 1=on.	**1**	
Ischaemic heart disease	0=off, 1=on.	**1**	Related to PA, air quality
Ischaemic stroke	0=off, 1=on.	**1**	Related to PA, air quality
Diabetes	0=off, 1=on.	**1**	Related to PA, air quality
Colorectal cancer	0=off, 1=on.	**1**	Related to PA
Breast cancer	0=off, 1=on.	**1**	Related to PA
COPD	0=off, 1=on.	**1**	Related to PA
Lung cancer	0=off, 1=on.	**1**	Related to PA
new Depression	0=off, 1=on.	**1**	Related to PA
new Anxiety	0=off, 1=on.	**1**	Related to PA
new Osteoarthritis	0=off, 1=on.	**0**	Related to PA
new Low back pain	0=off, 1=on.	**0**	Related to PA
new Sub-arachnoidal heamorrhage	0=off, 1=on.	**1**	Related to air quality
new Intra-cerebral heamorrhage	0=off, 1=on.	**1**	Related to air quality
new Lower respiratory tract infection	0=off, 1=on.	**1**	Related to air quality
PA-related all-cause mortality	0=off, 1=Woodcock, 2=accelerometry, 3=Arem, 4=Woodcock linear	**2**	Bypasses mortality via diseases
$PM_{2.5}$-related all-cause mortality	0=off, 1=on	**0**	*Not currently active in model*

Intervention decay	
Year	Effect over time
1	**100%**
2	**0%**
3	**0%**
4	**0%**
5	**0%**
6	**0%**
7	**0%**
8	**0%**
9	**0%**
10	**0%**
11	**0%**
12	**0%**
13	**0%**
14	**0%**
15	**0%**
16	**0%**
17	**0%**
18	**0%**
19	**0%**
20	**0%**
21	**0%**
22	**0%**
23	**0%**
24	**0%**
25	**0%**
26	**0%**
27	**0%**
28	**0%**
29	**0%**
30	**0%**
31	**0%**
32	**0%**
33	**0%**
34	**0%**
35	**0%**
36	**0%**
37	**0%**
38	**0%**
39	**0%**
40	**0%**
41	**0%**
42	**0%**
43	**0%**
44	**0%**
45	**0%**
46	**0%**
47	**0%**
48	**0%**
49	**0%**
50	**0%**
51	**0%**
52	**0%**
53	**0%**
54	**0%**

Note: The years in the 'intervention decay' column continue down to 101, when all cohorts have died.

Figure 13 shows a screen print of this 'dashboard'. The settings used in the scenario we used as reference case (walking) are shown on the dashboard in

Dashboard

Before running the model, enter/check the parameter settings
Values in bold red font denotes are used in the model
Bold black for making selections

Intervention				
Change can be made via mode share (population level) or change in number of trips per week for a number of persons (across ages and sex, as per current mode share - 'he who has, shall be given').				
Choose '1' for mode share, or '2' for change in number of trips for a number of persons			**2**	
Transport mode		**Baseline mode share**	**Intervention n**	**Change in number of trips pppw**
Private vehicle (driver/passenger)		71.9%	71.9%	-5
Walking		14.6%	**14.6%**	**5**
Cycle		1.9%	**1.9%**	**0**
Public Transport		11.5%	**11.5%**	**0**
Number of persons involved				**10,000**
Target population				
Age	Start		**20**	
	End		**100**	
Impact by activity level	0 = All categories (default)		**0**	
	1 = Inactive only			
	2 = Low active only			
	3 = Moderately active only			
	4 = Highly active only			
Economic values				
Discounting	Costs		**7%**	
	Health effects		**3%**	
Value of Life				
	Value of HALY1		**$ 213,000**	OBPR guidance note, Aug 2019
	Value of a Statistical Life 2		**$ 7,713,394**	TfNSW 2018 Guidelines econ appraisal
	Value of HALY2 (3% disc conversion TfNSWVSL to VSLY)		$ 333,700	
	Value of HALY3 (7% disc conversion OBPRVSL to VSLY)		$ 367,545	
Risk factors, diseases, mortality				
Physical activity	0=off, 1=on.		**1**	
% of added travel PA NOT offset by reduction in other PA			**100%**	(100% = all travel PA is additional)
Air quality ($PM_{2.5}$) - ambient	0=off, 1=on.		**1**	
Air quality ($PM_{2.5}$) - traffic participants	0=off, 1=on.		**1**	
Road traffic injury	0=off, 1=on.		**1**	
Ischaemic heart disease	0=off, 1=on.		**1**	Related to PA, air quality
Ischaemic stroke	0=off, 1=on.		**1**	Related to PA, air quality
Diabetes	0=off, 1=on.		**1**	Related to PA, air quality
Colorectal cancer	0=off, 1=on.		**1**	Related to PA
Breast cancer	0=off, 1=on.		**1**	Related to PA
COPD	0=off, 1=on.		**1**	Related to PA
Lung cancer	0=off, 1=on.		**1**	Related to PA
new Depression	0=off, 1=on.		**1**	Related to PA
new Anxiety	0=off, 1=on.		**1**	Related to PA
new Osteoarthritis	0=off, 1=on.		**0**	Related to PA
new Low back pain	0=off, 1=on.		**0**	Related to PA
new Sub-arachnoidal heamorrhage	0=off, 1=on.		**1**	Related to air quality
new Intra-cerebral heamorrhage	0=off, 1=on.		**1**	Related to air quality
new Lower respiratory tract infection	0=off, 1=on.		**1**	Related to air quality
PA-related all-cause mortality	0=off, 1=Woodcock, 2=accelerometry, 3=Arem, 4=Woodcock linear		**2**	Bypasses mortality via diseases
$PM_{2.5}$-related all-cause mortality	0=off, 1=on		**0**	*Not currently active in model*

Intervention decay	
Year	Effect over time
1	**100%**
2	**0%**
3	**0%**
4	**0%**
5	**0%**
6	**0%**
7	**0%**
8	**0%**
9	**0%**
10	**0%**
11	**0%**
12	**0%**
13	**0%**
14	**0%**
15	**0%**
16	**0%**
17	**0%**
18	**0%**
19	**0%**
20	**0%**
21	**0%**
22	**0%**
23	**0%**
24	**0%**
25	**0%**
26	**0%**
27	**0%**
28	**0%**
29	**0%**
30	**0%**
31	**0%**
32	**0%**
33	**0%**
34	**0%**
35	**0%**
36	**0%**
37	**0%**
38	**0%**
39	**0%**
40	**0%**
41	**0%**
42	**0%**
43	**0%**
44	**0%**
45	**0%**
46	**0%**
47	**0%**
48	**0%**
49	**0%**
50	**0%**
51	**0%**
52	**0%**
53	**0%**
54	**0%**

Note: The years in the 'intervention decay' column continue down to 101, when all cohorts have died.
Figure 13.

Model assumptions for the reference case:

- Shift towards walking, cycling and public transport go at the expense of the mode share for private vehicles (cars) [6].
- New walking trips replace car trips of 0-2km, average ~1.2km depending on age and gender [83]
- Bicycle trips replace car trips of 2-5km, average ~3.3km [83]
- Public transport trips replace car trips of 6-16km, average ~9km [83]
- The reference case assumes that 10,000 persons switch 5 trips per week from car to active transport. This is meant as an example. The number of people does not affect the per-km values. The assumption can be modified to suit the needs for a specific analysis.

- New users of active transport are like the current users of that active transport mode in terms of age and gender (which means that no infrastructure would get people over age 70 on a bike, for instance) (authors' assumption).
- Within each age-sex group, all activity levels participate equally when the mode share of AT is expanded (not only those already active, for instance). i.e., the amount of physical activity added by increased use of AT is the same for those already active, as for those currently inactive (authors' assumption). Instead of moving a proportion of the population from one activity category to the next higher, the model keeps everyone in the same group but lowers the risk of disease for that whole group, based on the shift in the average PA level in that group. In technical terms, this is the 'relative risk shift' method described in Barendregt and Veerman [84]
- We assumed an average speed of 4.48 km/h for walking, the lower end of the range in Ainsworth's compendium of physical activities, to relate time and distance, and 16 km/h for cycling. Varying this would have a modest impact on outcomes.
- Target population age 20 to 100 years. Younger generations can be included, but health impacts are absent until adult years (chronic disease is rare, and the current model does not include road traffic injury risks for children).
- All risk factor effects included: physical activity, ambient air quality, traffic participants' air quality, road traffic injury
- Physical activity from active transport is in addition to existing levels of PA (see Appendix P 11.Appendix Pfor more detail)
- All diseases included, except osteoarthritis and low back pain for which the evidence was not deemed strong enough
- Direct effect of physical activity on all-cause mortality, based on accelerometry-measured PA (option 2 in dashboard)
- Intervention with 1 year of full impact, and no effect thereafter, to give an estimate of the expected annual impact.
- Costs are discounted at 7% while health effects are discounted at 3%
- The value of a healthy life year is based on OBPR guidelines (see chapter 6)

In a sensitivity analyses, we systematically vary these settings and examine the impact this has on the results.

Dashboard

Before running the model, enter/check the parameter settings

Values in bold red font denotes are used in the model

Bold black for making selections

Intervention					**Intervention decay**	
Change can be made via mode share (population level) or change in number of trips per week for a number of persons (across ages and sex, as per current mode share - 'he who has, shall be given').					Year	Effect over time
Choose '1' for mode share, or '2' for change in number of trips for a number of persons				**2**	1	100%
					2	0%
Transport mode		**Baseline mode share**	**Intervention n**	**Change in number of trips pppw**	3	0%
Private vehicle (driver/passenger)		71.9%	71.9%	-5	4	0%
Walking		14.6%	**14.6%**	**5**	5	0%
Cycle		1.9%	**1.9%**	**0**	6	0%
Public Transport		11.5%	**11.5%**	**0**	7	0%
Number of persons involved				**10,000**	8	0%
					9	0%
Target population					10	0%
Age	Start		**20**		11	0%
	End		**100**		12	0%
Impact by activity level	0 =	All categories (default)	**0**		13	0%
	1 =	Inactive only			14	0%
	2 =	Low active only			15	0%
	3 =	Moderately active only			16	0%
	4 =	Highly active only			17	0%
					18	0%
Economic values					19	0%
Discounting	Costs		**7%**		20	0%
	Health effects		**3%**		21	0%
					22	0%
Value of Life					23	0%
	Value of HALY1		**$ 213,000**	OBPR guidance note, Aug 2019	24	0%
	Value of a Statistical Life 2		**$ 7,713,394**	TfNSW 2018 Guidelines econ appraisal	25	0%
	Value of HALY2 (3% disc conversion TfNSWVSL to VSLY)		$ 333,700		26	0%
	Value of HALY3 (7% disc conversion OBPRVSL to VSLY)		$ 367,545		27	0%
					28	0%
Risk factors, diseases, mortality					29	0%
					30	0%
Physical activity		0=off, 1=on.	**1**		31	0%
% of added travel PA NOT offset by reduction in other PA			**100%**	(100% = all travel PA is additional)	32	0%
Air quality ($PM_{2.5}$) - ambient		0=off, 1=on.	**1**		33	0%
Air quality ($PM_{2.5}$) - traffic participants		0=off, 1=on.	**1**		34	0%
Road traffic injury		0=off, 1=on.	**1**		35	0%
					36	0%
Ischaemic heart disease		0=off, 1=on.	**1**	Related to PA, air quality	37	0%
Ischaemic stroke		0=off, 1=on.	**1**	Related to PA, air quality	38	0%
Diabetes		0=off, 1=on.	**1**	Related to PA, air quality	39	0%
Colorectal cancer		0=off, 1=on.	**1**	Related to PA	40	0%
Breast cancer		0=off, 1=on.	**1**	Related to PA	41	0%
COPD		0=off, 1=on.	**1**	Related to PA	42	0%
Lung cancer		0=off, 1=on.	**1**	Related to PA	43	0%
new Depression		0=off, 1=on.	**1**	Related to PA	44	0%
new Anxiety		0=off, 1=on.	**1**	Related to PA	45	0%
new Osteoarthritis		0=off, 1=on.	**0**	Related to PA	46	0%
new Low back pain		0=off, 1=on.	**0**	Related to PA	47	0%
new Sub-arachnoidal heamorrhage		0=off, 1=on.	**1**	Related to air quality	48	0%
new Intra-cerebral heamorrhage		0=off, 1=on.	**1**	Related to air quality	49	0%
new Lower respiratory tract infection		0=off, 1=on.	**1**	Related to air quality	50	0%
					51	0%
PA-related all-cause mortality		0=off, 1=Woodcock, 2=accelerometry, 3=Arem, 4=Woodcock linear	**2**	Bypasses mortality via diseases	52	0%
$PM_{2.5}$-related all-cause mortality		0=off, 1=on	**0**	*Not currently active in model*	53	0%
					54	0%

Note: The years in the 'intervention decay' column continue down to 101, when all cohorts have died.

Figure 13 Overview of dashboard

The top of the dashboard allows to choose between modelling a change in the mode share of trips across the NSW population, or a change in the number of trips per week for each of the three active transport modes (walking, cycling, public transport). A table shows the mode share of trips as it is assumed to be at present and invites the analyst to define the mode share after intervention (option 1). The right-hand side of the panel permits to select the effect over time ('decay'), where 100% is the full effect and 0% no effect. The reference case arbitrarily assumes full effect for one year, with no effect thereafter. This can be varied based on information from specific projects.

The dashboard allows to include all activity categories (reference case) or produce per-km values separately for the inactive, low active, moderately active or highly active sections of the population.

The discount rate can be varied separately for economic outcomes and for health outcomes.

As discussed above, in the standard settings, three different values for a health-adjusted life year are applied, but these can be replaced by any value the analyst is interested in.

The risk factors, PA, ambient and traffic participant air quality, road traffic injury, can be switched on or off (in the reference case, all are 'on'). For PA, a cell allows to indicate the proportion of travel-related physical activity that is deemed to be the change net of any offsetting changes in physical activity in other domains (leisure, housework, work-related). This allows for the possibility that as people are more active for transport, they might be less active in other domains because, for example, travel by active transport takes up time that otherwise might be spent in leisure-time PA, or because people judge they have already achieved their daily dose of exercise and therefore skip the gym.

We did a rapid systematic search of the literature for empirical evidence of this displacement effect (Appendix P). A study in the UK published in 2015 ('Changes in active commuting and changes in physical activity in adults: a cohort study') found that 'changes in active commuting were associated with commensurate changes in total self-reported physical activity and we found no compensatory changes in self-reported recreational physical activity' [85]. A later study by the same group used different methods and data and found that 'Compared to those not undertaking active travel, those who did active travel reported 11 min more in leisure MVPA and 18 min less in screen time per day, and reported lower sleep [86]. This points to the possibility that rather than reducing physical activity in other domains, increased active travel could be associated with increased activity in leisure time, and hence the authors conclude that "overall, active travel was associated with a broadly health-promoting composition of time across multiple behavioural domains, which supports the public health case for active travel". A US study found that "results, albeit cross-sectional, are not supportive of the hypothesis that active transportation displaces time spent in leisure activity and rather suggest co-occurrence" [87]. Evidence that does not directly address this issue does support the notion that active travel is associated with better physical fitness: "Majority [sic] of the investigations on young ages and adults have shown positive effects or relationships between active commuting and several attributes of physical fitness" [88]. A study in Canada found that higher neighbourhood walkability was associated with decreased prevalence of overweight and obesity and decreased incidence of diabetes between 2001 and 2012 [10]. Overweight, obesity and diabetes are related to physical activity, so this provides indirect evidence of a net positive contribution to levels of PA by active transport. In sum, a search of the peer-reviewed published literature does not support a displacement effect of active travel and suggests that the physical activity obtained in active travel is additional to leisure time and other forms of physical activity. To corroborate this finding, we recommend further research through a comprehensive systematic review of current literature.

Diseases can also be switched on or off. In the reference case, all are 'on' except for OA and LBP, for which the current evidence of a causal relationship with PA was deemed too weak. The model does not have individuals as the unit of analysis, but instead works with populations (1-year age/sex groups), of which a proportion has a decrement in quality of life due to a disease. That proportion changes slightly as the prevalence of the disease changes. Changes in mortality, likewise, are calculated in rates for each disease. Mortality rates can validly be added, and on the basis of the overall rates, 1-year probabilities are calculated. The life table structure then avoids double counting.

For PA, ACM can be switched to '1', in which case the model uses the relative risks by Woodcock and colleagues [64] to link changes in PA to changes in mortality, without lag time and bypassing the

mortality via the diseases in the model (effects on morbidity and quality of life are unaffected). Switching ACM to '2' applied accelerometry-based relative risks [30], and this was selected for the reference case analysis because it is based on the strongest study design (systematic review and meta-analysis of good quality cohort studies with sufficient follow-up and the exposure is measured using gold-standard methods). Options '3' and '4' refer to the use of the results from a large cohort study [65] and a variant of the Woodcock study [64] in which we replaced the original dose-response relationship modelled using a power function, with linear extrapolation between the values given. The power function is very steep at the extreme low end of the distribution (no PA), which may lead to overestimation of the impact of shifts in the 'inactive' category; the linear interpolation avoids this.

Table 15 presents results for the reference case and Table 16 those for a range of scenarios in the sensitivity analysis, for selected outcomes. The findings will be discussed below.

Table 15 Main results – reference case

Health discounted 3% Costs discounted 7%	Walking	Cycling (on road)	Cycling (off road)	Walking associated with public transport use
	$ per km	$ per km	$ per km	$ per km
Reference case	5.42 (5.11 - 5.78)	1.47 (1.38 - 1.59)	1.58 (1.48 - 1.69)	3.53 (3.29 - 3.79)
Physical activity	5.61 (5.28 - 5.94)	1.58 (1.48 - 1.68)	1.58 (1.48 - 1.68)	3.48 (3.24 - 3.70)
Air quality ($PM_{2.5}$) - ambient	0.01 (0.01 - 0.02)	0.00 (0.00 - 0.00)	0.00 (0.00 - 0.00)	0.00 (0.00 - 0.00)
Air quality ($PM_{2.5}$) - traffic participants	-0.02 (-0.03 - -0.01)	-0.02 (-0.02 - -0.01)	#	-0.01 (-0.02 - -0.01)
Road traffic injury	-0.18 (-0.18 - -0.18)	-0.08 (-0.08 - -0.08)	#	0.06 (0.06 - 0.07)

Notes: The reference case assumes 10,000 travellers switch 5 trips per week from car to active transport. Exposures included are physical activity, air quality (PM2.5, ambient and traffic participants), road trauma. The conversion of VSL to VSLY is discounted at 3% and VSLY is $213,000 (OBPR). *For the off-road cycling scenario, RTI and air pollution for traffic participants switched off, leaving physical activity and ambient air pollution effects. #walking and cycling to and from public transport transit stops

The reference case puts the economic value of a km walking at $5.42 (95% uncertainty range $5.11 to $5.78). An additional km cycled results in $1.47 ($1.38 to $1.59) in health benefits (Table 15). An increase in public transport use is associated with health benefits of $3.53 ($3.29 to $3.79). These are marginal benefits – they represent the effect of adding physical activity to existing average activity levels in the NSW population. The difference in these values depends on various factors. The lower value for cycling compared to walking is partly explained by the greater efficiency of cycling. The age of the people who walk, cycle or use public transport also has a large influence on the size of the health benefits, as we shall see below. The absolute benefits of extra physical activity are much greater in old age compared to young age. Cyclists tend to be relatively young, and this reduces the value of a km cycled compared to the other modes, which are used more by older travellers. The output of the model 'translates' all physical activity to both walking and cycling (regardless of whether the scenario involved replacing car trips with walking, cycling or public transport), making it possible to create output that removes the part of this difference that is created by age. (It then gives results for either the age mix of walkers, cyclists, or users of public transport.) Removing the effect of age leaves a ratio of 1.54, which agrees with current TfNSW assumptions regarding the relative intensity of effort per

km of cycling compared to walking. (In other words, if new cyclists had the same age distribution as current walkers, the value of a km cycling would be about 1.5 times higher than estimated above.)

The impact of a shift to off-road cycling has been estimated running the analysis for cycling with the effects for traffic injury and air pollution exposure of traffic participants switched off. This results in a value of $1.58 ($1.48 - $1.69) per km (Table 15). It requires the assumption that the off-road infrastructure is perfectly safe, and its users are exposed to background air quality rather than the worse air quality when sharing the road with motorised traffic, and that the risk and exposure estimates in the reference scenario are based on on-road walking and cycling. The impact of shifting existing cycling trips from on-road to off-road can be estimated by taking the difference between the outcomes of the two scenarios: one with a specified increase in walking or cycling with the above two switches off (for off-road) and one with them on (on road). This results in a value of $0.11 per km for a shift of these trips from on-road to off-road for avoided risks of RTI and air pollution exposure.

Nearly all the benefits are from increases in PA, which is valued at $5.42 per km for walking and $1.47 per km for cycling. The positive effects of improvements in ambient air quality are negligible and offset by almost equally small negative effects of active transport participants inhaling more pollutants. Road traffic injuries and deaths are estimated to result in losses of $0.18 (reported as – 0.18) per km worth of health and lifetime if car trips are replaced by walking, and $0.08 per km for cycling. This effect occurs because on a per km basis, walking and cycling are riskier than car transport. In contrast, replacing car trips with public transport delivers a net benefit of $0.06 per km, as public transport is safer than private car use, and also because taking cars off the road reduces the risks faced by other traffic participants.

Table 16 shows the results from the sensitivity analysis.

Table 16 Main results – sensitivity analysis

	Walking	Cycling (on road)	Cycling (off road)	Walking associated with public transport use
	$ per km	$ per km	$ per km	$ per km
Value of a healthy life year				
OBPR (VSLY at $213,000 from conversion of VSL to VLY based on a private time preference discount rate of 3%) - **reference case**	**5.42**	**1.47**	**1.58**	**3.53**
ATAP/TfNSW (VSL at $7,713,393 in 2019 $)	8.48	2.30	2.47	5.50
7 % discounting of OBPR VSL to Value of HALY	9.33	2.54	2.72	6.06
Discount rates				
0% for costs, 0% for health outcomes	7.86	2.31	2.47	5.38
3% for costs, 3% for health outcomes	5.48	1.48	1.58	3.53
7% for costs, 7% for health outcomes	3.76	0.95	1.05	2.37
7% for costs, 5% for health outcomes	4.46	1.16	1.26	2.84
10% for costs, 10% for health outcomes	3.02	0.75	0.83	1.92
Health care costs				
Excl. health care costs ($HALY per km without HCC)	5.39	1.47	1.57	3.49
Risk factors				
Physical activity -100% additional to other PA	5.61	1.58	1.58	3.48
Physical activity - 80% additional to other PA	4.89	1.39	1.39	2.96
Physical activity - 50% additional to other PA	3.23	1.07	1.07	2.03
Diseases				
'Established' diseases only	0.51			
Diseases as reference case, all-cause mortality off	0.76			
Depression, via physical activity	0.12			
Anxiety, via physical activity	0.14			
Osteoarthritis, via physical activity	0.74			
Low back pain, via physical activity	0.21			
Lower respiratory infections, via PM2.5	-0.01			
Intracerebral haemorrhage, via PM2.5	0.00			
Subarachnoid haemorrhage, via PM2.5	0.00			
All-cause mortality Woodcock's power fit	1.61			
All-cause mortality Woodcock, linear fit	1.48			
All-cause mortality, Arem et al 2015	2.06			
Intervention				
Large intervention: doubling of active transport, incl. public transport	5.74			
Small intervention: 1% rel. increase in active & public transport	11.88			
Mode share walking +10 percentage points	7.32			

Mode share cycling +10 percentage points	4.63			
Mode share public transport +10 percentage points	6.78			
Intervention effect lasts lifelong	5.56			
Physical activity level of new active transport participants				
Inactive	8.93			
Low active	8.26			
Moderately active	2.06			
Highly active	0.05			
Age				
Age 20-29 years	0.90			
Age 30-39 years	1.81			
Age 40-49 years	3.22			
Age 50-59 years	7.44			
Age 60-69 years	13.77			
Age 70+ years	26.92			

Notes: Unless otherwise indicated, reference case settings (see Figure 14 and text above) were used. Established diseases are those that were included in the model used in Phase 1 of this project (and the 2015 Global Burden of Disease): breast cancer, colon cancer, type 2 diabetes, ischemic stroke, ischemic heart disease, chronic obstructive pulmonary disease and lung cancer, and road traffic injuries and deaths. *The impact of cycling on separate, off-road bikeways was estimated by switching off the impacts of RTI and air pollution exposure for travellers. Not all scenarios for cycling and public transport were run due to time constraints.

When using the TfNSW value of a statistical life, the value of a km walking is $8.48, a bit over $3 less than the $5.42 when using the OBPR value. However, these numbers are sensitive to several of the parameter settings, as is the results of the sensitivity analysis show.

Health care costs make little difference to the results. In the walking reference scenario, they contribute about $0.04 per km walking, and $0.01 per km for cycling (Table 16).

In the sensitivity analysis of this study, we explored various combinations of discount rates for economic outcomes and for health outcomes. The discount rate for costs does not have much influence on the results, as the health care costs make up a very modest component of the overall values. The rate at which future health outcomes are discounted has greater impact. With 0% discounting of health effects and costs, the value of a km walking is $7.86 while a km cycling on road is $2.31. An additional km cycling off road was valued at $ 2.47 and a km walking with public transport use was $5.38. A discount rate of 10% for both costs and health results in a value of $3.02 for walking and $0.75 per km for cycling. A uniform discount rate of 7% puts this value at $3.76, which is just over 30% lower than the reference case value of $5.42, where health is discounted at 3%.

The results show that if extra physical activity from active transport were offset by less physical activity in other domains of life (leisure, home, work), then the value per km would go down. However, as discussed above (page 45), the evidence suggests that these assumptions are too pessimistic.

The newly added health outcomes explain most of the value of a km walked, explaining over 90% of the value of a km walked ($4.92 of the $5.42). By far most of this is due to the adding of a direct link between physical activity and the risk of death. The inclusion of a direct effect of PA on overall mortality raises the value of a km walked about three- to five-fold, from $0.76 to $5.42. The accelerometer-based values that were used in the main analysis give much larger effects compared to values in which physical activity was assessed in surveys.

Compared to only including the diseases in the Phase 1 model, depression adds $0.12 and anxiety $0.14 per km walking. If switched on, osteoarthritis adds $0.74 and low back pain $0.21. The values for osteoarthritis and (to a lesser extent) low back pain are based on limited evidence, however, and were not included in the main analysis. The addition of lower respiratory infections and two types of stroke make very moderate differences; they have a weak link with air quality, which itself is not greatly influenced by the modest changes in traffic modes that were modelled. Lower respiratory infections even show a slightly negative value, which may be due to additional cases of lower respiratory infections causing deaths in the extra year of life added by increased levels of physical activity.

Doubling the use of active transport produces a value of the average extra km of walking at $5.74, whereas a small increase in the use of active transport (1% extra) results in a value of $11.88. The high value for smaller interventions can be explained by the risk curve for PA being very steep at the lower end, so a few minutes of walking per week result in a comparatively large reduction in the risk of disease. This applies especially to the 'established' diseases, which are modelled with a power function. For the 'new' diseases we took a different approach in which linear interpolations between the mid-values of the activity categories determine the level of risk.

This effect is also seen when modelling the various activity groups separately. The value for the inactive and low active are around $8.93 per km walking, whereas in persons who are already highly active, the increased risks of road traffic injury or death and the small additional benefit from PA on non-communicable disease and mortality balance out.

As observed in Phase 1, age also has a very large impact on the size of the benefits of active transport. The value of a km walking rises from $0.90 for those in their twenties, to about $26.92 for those over 70. This is because the risk of the chronic diseases in the model is high in old age, and to a lesser extent because the discounting favours immediate gains over those decades into the future.

In conclusion, this analysis suggests values for the health-benefits of active transport that are higher than most previous estimates. This is mainly due to the inclusion of a direct link with mortality, especially if the PA-mortality relationship is based on studies in which PA is measured with accelerometry. The inclusion of mental health outcomes also contributes considerably. The evidence that physical activity can prevent musculoskeletal conditions was judged not strong enough for inclusion in the main analysis, but if further evidence confirms a causal relationship, low back pain and osteoarthritis could potentially add considerable value. The results vary systematically and substantially with age and activity level, with great benefits for the old and the inactive and much lower gains for the young, and almost none for the segment of the population that already achieves high levels of PA. The choice of the value for a healthy life year and discount rates for health effects and costs also has a large impact.

8. How the NSW Active Transport Health Model complies with the NSW Government Guide to Cost Benefit Analysis, NSW Treasury

In NSW, major infrastructure measures have to undergo cost benefit analysis. NSW Treasury sets out guidelines for cost benefit analysis, detailed in the NSW government guide to cost benefit analysis [2]. This involves nine consecutive steps:

1. Stating the objectives
2. Define the reference case and develop options
3. Identify and forecast costs and benefits
4. Value the costs and benefits
5. Identify qualitative factors and distributional impacts
6. Assess risks and test sensitivities
7. Assess net benefits
8. Report the results
9. Undertake post evaluation

Although the current project only focussed on the health benefits of active transport, which are typically just one component among many in the assessment of major infrastructure measures, the project followed Treasury NSW guidelines for CBA where applicable. Table 17 details the key steps recommended by Treasury NSW and how these have been addressed in the current study if applicable.

Table 17 Cost Benefit Analysis Steps and the NSW Active Transport Health Model

Number	Step	Comment
1	Stating the objectives	Objectives for the overall project will be stated within the wider CBA. The objectives of this project were to develop a model to estimate the cost benefits of active transport in NSW and to derive cost values through application of the model.
2	Define the reference case and develop options	The model is run using a reference case presented in section 0. The definition of the reference case needs to be discussed and agreed upon for a final version of these analyses.
3	Identify and forecast costs and benefits	The model forecasts cost and benefits through modelling a change in the mode share of trips across the NSW population, or a change in the number of trips per week for each of the three active transport modes (walking, cycling, public transport).
4	Value the costs and benefits	The model produces results summarised in 'health-adjusted life years' (HALYs) that can be converted into monetary values. The model incorporates the monetary values most commonly used in the Australian context: the ATAP/TfNSW 'Inclusive Willingness to Pay' value, and the OBPR/Abelson value. Both are derived from the 'value of a statistical life' (VSL).
5	Identify qualitative factors and distributional impacts	N/A
6	Assess risks and test sensitivities	N/A part of wider CBA
7	Assess net benefits	The model calculates benefits of active transport not net benefits of infrastructure measure

8	Report the results	The results are reported in the NSW Active Transport Health Model output (see section 0). These can then be integrated into CBA.
9	Undertake post evaluation	N/A

The CBA guidelines also identify key conditions that need to be taken into account within CBAs including discounting, life of project, net benefits, sensitivity analysis, avoidance of double counting, NSW community as referent group. The conditions have been incorporated into the model.

Please note that in the reference case in this report, a discount rate of 3% is used for health outcomes, which is consistent with the social time preference rate used by Treasury. Treasury recommends the use of the 7% discount rate that is intended to represent the opportunity cost of capital (see Supplementary file 1 and section 6.4.2), which was applied in a sensitivity analysis (Table 16). We recommend that in estimating the value of health-related benefits from active transport in future years, analysts of business cases apply a 3% discount rate there as well.

As explained in Chapters 5 and 7, double counting has been avoided by the use of a lifetable to integrate the shifts in the epidemiology of the various health conditions in the model, and by the automatic switching off of disease specific mortality when a direct link between physical activity and mortality of any cause is switched on. Furthermore, the valuations of HALY do not include the value of the changes in health care costs that are also included in the model.

9. Application and integration of the NSW Active Transport Health Model

The NSW Active Transport Health Model can be applied in full, as a complete model, or in part, where the output parameters are applied separately to cost health benefits accrued per km walked or cycled. During phase one of the project, a range of opportunities to embed the model in the cost – benefit processes of agencies across NSW were identified. These include the Infrastructure NSW "Business Case Toolkit" [89] and the TfNSW "Transport for NSW Economic Parameter Values" [90]. The health benefit cost/kilometre could potentially be incorporated into the Outcome Values Data Bank.

Options for how the NSW Active Transport Health Model can be applied NSW and integrated across government agency portfolios will require further discussion with the relevant organisations directly and collectively through, for example, the NSW Government Cost Benefit Analysis Best Practice Working Group.

Potential use of the NSW Active Transport Health Model in infrastructure planning by TfNSW and how the parameters values generated by the model align with those currently recommended by TfNSW will be discussed in more detail below.

The model could also be used at a local or precinct level to inform planning and guide evaluation of projects such as healthy streets initiatives [91]. Moreover, it could be used to estimate health impacts and health-related economic benefits of initiatives targeted at areas characterised by specific demographic characteristics of the local population which benefits from use of tailored values. The model could also be integrated into health impact assessments of proposals that involve active transport components (e.g., health infrastructure developments).

Use of the full model would require training and relevant skills and expertise such as epidemiology. This could be accommodated through housing the model within the NSW Ministry of Health. In a future phase of the project, we plan to provide a user guide on how to use the model. We also plan to train key staff in the use.

9.1 Integration of the NSW Active Transport Health Model in TfNSW practice

Transport for NSW considers active transport as part of the planning of transport services in NSW [92]. The principles and guidelines for the economic appraisal of transport projects and initiatives are outlined in the TfNSW document "Principles and Guidelines for Economic Appraisal of Transport Investment and Initiatives" [92]. The recommended economic parameter values for the cost benefit analysis of transport appraisals are detailed in the accompanying document "Transport for NSW Economic Parameter Values" [90]. These two documents provide best-practice approaches and economic parameter values and support the consistent application of cost-benefit analysis (CBA) across the NSW Transport cluster. They also include specific guidance on active transport. The health benefits of active transport (walking and cycling) due to increased physical activity as well as the potential disbenefits due to road transport injuries and air pollution are specified as separate $ per km values in these guidance documents for use in cost benefit analysis. The NSW Active Transport Health Model considers the joint health effects associated with the exposures physical activity, air pollution and road transport as a default, but values for each individual exposure can also be calculated separately. How these align with those currently recommended by TfNSW will be outlined below. Additional steps to account for non-health benefits are required for a full cost-benefit analysis.

9.1.1 Physical activity

The NSW Active Transport Health Model calculates and values the health benefits of increased physical activity due to active transport for the NSW population. The model considers the most up to date evidence on health outcomes associated with physical activity, and the underlying mathematical model is currently considered to be the best practice method to quantify the health benefits of active transport [93]. To determine how the parameters will be used in practice, we hope to pilot this on a few practical cases to identify the best use for this method, with the involvement of Transport for NSW, Treasury, and other relevant departments. The $ per km values generated by the model for physical activity can be readily adopted by TfNSW in their guidelines for economic appraisal of transport investment to cost the health benefits of active transport (Table 18).

The per-km values represent the net present value (discounted) of future health benefits that follow from one year of extra physical activity. The physical activity-related benefits in future years will have to be discounted at the rate that is applied to health outcomes (in this report, a rate of 3% is used for the reference case [Chapter 6]).

In CBAs, the model results should be applied as cumulative benefits based on the behavioural change timeframes of the relevant intervention. For example, for a project that replaces car trips with walking trips for 2 years, the CBA will calculate the total number of additional kms walked over year 1 multiplied by the per-km value ($5.42 in the reference case), plus the multiplication of the total number of additional kms walked over year 2 with $5.42, to which the discount rate for health outcomes should be applied.

9.1.2 Road transport injury

The NSW Active Transport Health Model quantifies the injury costs (quality and duration of life lost, and health care costs) for walking and cycling. The model calculates the injury costs based on the additional km walked and cycled due to uptake of active transport relative to the current injury rates. Based on international evidence, these calculations take into consideration a safety in numbers effect whereby increases in pedestrian and cyclist numbers results in a less than proportional increase in the number of injuries [39].

In comparison, TfNSW recommends the inclusion of the crash costs in cost benefit analyses of infrastructure measures that consider active transport. Crash costs include the costs incurred by injury and property damage. TfNSW recommends that road safety benefits be estimated based on the Inclusive Willingness-to-Pay (WTP) approach. Different cost values are used depending on crash outcome (injury or deaths) and severity of the injury. In comparison, the NSW Active Transport Health Model only considers health related costs (i.e., the value of quality and duration of life lost, and health care costs).

Given these different approaches to calculating the crash costs associated with active transport there is a need to further explore how the road transport injury values from the NSW Active Transport Health Model could be used by TfNSW and whether changes to the NSW Active Transport Health Model are needed in future to harmonize methods.

9.1.3 Air pollution

The NSW Active Transport Health Model considers the adverse health impacts of air pollution (on those engaging in active transport as well as benefits to the wider population from less pollutants due to reduced motorised transport. Those participating in active transport have higher exposure to air pollution from road transport, but overall levels of air pollution from road transport decrease if more people switch from motorised to active transport. The health outcomes associated with air pollution considered in the NSW Transport Health Model are based on the most up to date evidence. The model uses $PM_{2.5}$ as a measure of air pollution is, this is derived from measuring stations in the greater Sydney area.

In comparison, TfNSW currently considers the negative health effects of increased exposure to air pollution in those who engage in active transport but not the health benefits of improved air quality due to reduced motorised transport. For application in metropolitan areas, the values for air pollution from the NSW Active Transport Health model could be adopted by TfNSW for use in cost benefits analysis. This will ensure that the benefits that accrue from improved air quality at population level are also considered in economic appraisals of active transport.

Table 18 Active transport parameters currently recommended by TfNSW* and values generated by the NSW Active Transport Health Model

	TfNSW		**NSW Active Transport Health Model**			
Cost / Benefit	**Cycling ($/km)**	**Walking ($/km)**	**Cycling on road ($/km)**	**Cycling off-road ($/km)**	**Walking ($/km)**	**Walking associated with public transport use ($/km)**
Health benefits #	**1.22**	**1.83**	**1.58**	**1.58**	**5.61**	**3.48**
Congestion cost savings	0.41	0.41	N/A	N/A	N/A	N/A
Vehicle operating cost savings	0.37	0.42	N/A	N/A	N/A	N/A
Accident cost$	0.24	0.12	-0.08	x	-0.18	0.06
Air pollution cycling / walking&	0.03	0.03	-0.02	x	-0.02	-0.01
Air pollution population level	N/A	N/A	0.00	0.00	0.01	0.00
GHG emissions	0.03	0.03	N/A	N/A	N/A	N/A
Noise	0.01	0.01	N/A	N/A	N/A	N/A
Water pollution	0.01	0.01	N/A	N/A	N/A	N/A
Nature and landscape	0.00	0.00	N/A	N/A	N/A	N/A
Urban separation	0.01	0.01	N/A	N/A	N/A	N/A
Roadway provision cost savings	0.04	0.04	N/A	N/A	N/A	N/A
Parking cost savings	0.01	0.01	N/A	N/A	N/A	N/A

* TfNSW values, recipients: Former car users and for health benefits former car and public transport users

Values from NSW Active Transport Model Health refer to physical activity only, not including air pollution and RTI.

x Accident cost and air quality effects for off-road cycling were assumed negligible in this analysis.

$ Injury costs only in the NSW active Transport Health model. Positive for public transport because PT is safe and takes cars off the road, which offsets the increased risk when walking to and from transit stops.

& The slight negative benefit for air pollution cycling/walking from NSW Active Transport Model Health is related to increased exposure to air pollution for traffic participants while walking, cycling or walking to and from transit stops

*Source: Transport for NSW Economic Parameter Values Evaluation & Assurance Group Finance & Investment Corporate Services.2019 [90].

10. References

1. Australian Transport Assessment and Planning (ATAP), *Australian Transport Asessment and Planning Guidelines M4 Active Travel*. 2015.

2. NSW Government Treasury, *Policy and Guidelines Paper: NSW Government Guide to Cost-Benefit Analysis (TPP17-03)*, Treasury, Editor. 2017, NSW Government www.treasury.nsw.gov.au/.

3. Booth, F.W., C.K. Roberts, and M.J. Laye, *Lack of exercise is a major cause of chronic diseases.* Comprehensive Physiology, 2012. **2**(2): p. 1143-1211.

4. Lee, I.M., et al., *Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy.* Lancet, 2012. **380**(9838): p. 219-29.

5. Giles-Corti, B., et al., *City planning and population health: a global challenge.* Lancet, 2016. **388**(10062): p. 2912-2924.

6. Australian Government, *Walking, riding and access to public transport. Supporting active travel in australian communities ministerial statement.*, Department of Infrastructure and Transport, Editor. 2013, Commonwealth of Australia https://www.infrastructure.gov.au/infrastructure/pab/active_transport/files/infra1874_mcu_active_travel_report_final.pdf.

7. Zapata-Diomedi, B., et al., *A shift from motorised travel to active transport: What are the potential health gains for an Australian city?* PloS one, 2017. **12**(10): p. e0184799-e0184799.

8. Woodcock, J., M. Givoni, and A.S. Morgan, *Health impact modelling of active travel visions for England and Wales using an Integrated Transport and Health Impact Modelling Tool (ITHIM).* PLoS One, 2013. **8**(1): p. e51462.

9. Cavill, N., et al., *The Health Economic Assessment Tool (HEAT) for walking and cycling: From evidence to advocacy on active transport.* Journal of Science and Medicine in Sport, 2012. **15**: p. S69.

10. Creatore, M.I., et al., *Association of Neighborhood Walkability With Change in Overweight, Obesity, and Diabetes.* JAMA, 2016. **315**(20): p. 2211-2220.

11. Frank, L.D., M.A. Andresen, and T.L. Schmid, *Obesity relationships with community design, physical activity, and time spent in cars.* American Journal of Preventive Medicine, 2004. **27**(2): p. 87-96.

12. King, A.C., et al., *Aging in neighborhoods differing in walkability and income: associations with physical activity and obesity in older adults.* Social science & medicine (1982), 2011. **73**(10): p. 1525-1533.

13. Ding, D., et al., *The economic burden of physical inactivity: a global analysis of major non-communicable diseases.* Lancet, 2016. **388**(10051): p. 1311-24.

14. World Health Organisation, *WHO Expert Meeting Methods and tools for assessing the health risks of air pollution at local, national and international level*. 2014.

15. Hamra, G.B., et al., *Outdoor particulate matter exposure and lung cancer: a systematic review and meta-analysis.* Environ Health Perspect, 2014. **122**(9): p. 906-11.

16. James, S.L., et al., *Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017.* The Lancet, 2018. **392**(10159): p. 1789-1858.

17. Institute for Health Metrics and Evaluation. *Global Burden of Disease Study 2017 (GBD 2017) Results*. 2018 [cited 2020 February]; Available from: http://ghdx.healthdata.org/gbd-results-tool.

18. Australian Bureau of Statistics. *Microdata: National Health Survey, 2017-18*. 2019 [cited 2020 April]; Available from: https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4324.0.55.001Explanatory%20Notes102017-18?OpenDocument.

19. NSW Planning Industry and Environment. *Air Quality Data*. 2018 [cited 2020 April]; Available from: https://www.dpie.nsw.gov.au/air-quality/search-for-and-download-air-quality-data.

20. Transport for New South Wales. *Centre for Road Safety Crashlink database* 2020 [cited 2020 March]; Available from: https://roadsafety.transport.nsw.gov.au/statistics/index.html.

21. Australian Institute of Health and Welfare. *Disease expenditure in Australia 2015- 2016*. 2019 [cited 2020 April]; Available from: https://www.aihw.gov.au/reports/health-welfare-expenditure/disease-expenditure-australia/data.

22. Global Burden of Disease, *Supplementary Appendix 1: GBD 2016 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2017; 390: 1345–422.* 2017.

23. World Cancer Research Fund. *Food, nutrition, physical activity and the prevention of cancer: a global perspective. The Third Expert Report*. 2018; Available from: https://www.wcrf.org/dietandcancer.

24. Australian Bureau of Statistics. *Estimated Resident Population By Single Year Of Age, New South Wales*. 2019 [cited 2020 February]; Available from: https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3101.0Sep%202019?OpenDocument#Data.

25. Barendregt, J.J., et al., *A generic model for the assessment of disease epidemiology: the computational basis of DisMod II.* Population Health Metrics, 2003. **1**(1): p. 4.

26. van Baal, P.H.M., et al., *Economic evaluation and the postponement of health care costs.* Health Economics, 2011. **20**(4): p. 432-445.

27. Australian Institute of Health and Welfare. *Health expenditure Australia 2017–18*. 2019 [cited 2020 August]; Available from: https://www.aihw.gov.au/reports/health-welfare-expenditure/health-expenditure-australia-2017-18/data.

28. NSW Health. *NSW Population Health Survey* 2017 [cited 2020 March]; Available from: https://www.health.nsw.gov.au/surveys/Pages/nsw-population-health-survey.aspx.

29. Ainsworth, B.E., et al., *Compendium of physical activities: an update of activity codes and MET intensities.* Med Sci Sports Exerc, 2000. **32**(9 Suppl): p. S498-504.

30. Ekelund, U., et al., *Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: systematic review and harmonised meta-analysis.* Bmj, 2019. **366**: p. l4570.

31. Johnston, F.H., et al., *Estimated global mortality attributable to smoke from landscape fires.* Environmental health perspectives, 2012. **120**(5): p. 695-701.

32. Chang, T.C.L., et al., *Major Source Contributions to Ambient PM2.5 and Exposures within the New South Wales Greater Metropolitan Region.* Atmosphere, 2019. **10**(3).

33. World Health Organization, *WHO Expert Meeting: Methods and tools for assessing the health risks of air pollution at local, national and international level*. 2014, WHO Regional Office for Europe: Copenhagen.

34. Rojas-Rueda, D., et al., *Replacing car trips by increasing bike and public transport in the greater Barcelona metropolitan area: A health impact assessment study.* Environment International, 2012. **49**: p. 100-109.

35. Tainio, M., et al., *Can air pollution negate the health benefits of cycling and walking?* Preventive Medicine, 2016. **87**: p. 233-236.

36. Maizlish, N., et al., *Health Cobenefits and Transportation-Related Reductions in Greenhouse Gas Emissions in the San Francisco Bay Area.* American Journal of Public Health, 2013. **103**(4): p. 703-709.

37. Australian Bureau of Statistics. *Survey of Motor Vehicle Use, Australia, 12 months ended 30 June 2018*. 2019 [cited 2020 April]; Available from: https://www.abs.gov.au/ausstats/abs@.nsf/mf/9208.0.

38. Transport Performance and Analytics. *Active Transport: Cycling*. 2019 [cited 2020 May]; Available from: https://opendata.transport.nsw.gov.au/dataset/active-transport-cycling.

39. Elvik, R. and R. Goel, *Safety-in-numbers: An updated meta-analysis of estimates.* Accident Analysis & Prevention, 2019. **129**: p. 136-147.

40. GBD 2017 Risk Factor Collaborators., *Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017.* Lancet, 2018. **392**(10159): p. 1923-1994.

41. Danaei, G., et al., *The preventable causes of death in the United States: comparative risk assessment of dietary, lifestyle, and metabolic risk factors.* PLoS Med, 2009. **6**(4): p. e1000058.

42. Kyu, H.H., et al., *Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: systematic review and dose-response meta-analysis for the Global Burden of Disease Study 2013.* BMJ, 2016. **354**: p. i3857.

43. Vos, T., et al., *Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013.* The Lancet, 2015. **386**(9995): p. 743-800.

44. Centre for Epidemiology and Evidence. *HealthStats NSW. Sydney: NSW Ministry of Health*. 2019 [cited 2020 June]; Available from: http://www.healthstats.nsw.gov.au/.

45. Webb, P., C. Bain, and A. Page, *Essential Epidemiology: An Introduction for Students and Health Professionals*. 4 ed. 2019: Cambridge University Press.

46. Rothman, K.J., S. Greenland, and T.L. Lash, *Modern Epidemiology, 3e (pb)*. 2008: Lippincott.

47. Mammen, G. and G. Faulkner, *Physical activity and the prevention of depression: a systematic review of prospective studies.* Am J Prev Med, 2013. **45**(5): p. 649-57.

48. Schuch, F.B., et al., *Physical activity and incident depression: A meta-analysis of prospective cohort studies.* American Journal of Psychiatry, 2018. **175**(7): p. 631-648.

49. McDowell, C.P., et al., *Physical Activity and Anxiety: A Systematic Review and Meta-analysis of Prospective Cohort Studies.* American Journal of Preventive Medicine, 2019. **57**(4): p. 545-556.

50. Schuch, F.B., et al., *Physical activity protects from incident anxiety: A meta-analysis of prospective cohort studies.* Depress Anxiety, 2019. **36**(9): p. 846-858.

51. Alzahrani, H., et al., *The association between physical activity and low back pain: a systematic review and meta-analysis of observational studies.* Scientific reports, 2019. **9**(1): p. 8244-8244.

52. Shiri, R. and K. Falah-Hassani, *Does leisure time physical activity protect against low back pain? Systematic review and meta-analysis of 36 prospective cohort studies.* British journal of sports medicine, 2017. **51**(19): p. 1410-1418.

53. Heneweer, H., et al., *Physical activity and low back pain: A systematic review of recent literature.* European Spine Journal, 2011. **20**(6): p. 826-845.

54. Hart, L.E., et al., *The relationship between exercise and osteoarthritis in the elderly.* Clin J Sport Med, 2008. **18**(6): p. 508-21.

55. Richmond, S.A., et al., *Are joint injury, sport activity, physical activity, obesity, or occupational activities predictors for osteoarthritis? A systematic review.* The Journal of orthopaedic and sports physical therapy, 2013. **43**(8): p. 515-B19.

56. White, D.K., et al., *Daily walking and the risk of incident functional limitation in knee osteoarthritis: an observational study.* Arthritis Care Res (Hoboken), 2014. **66**(9): p. 1328-36.

57. KylesConverter.com. *Convert Steps to Meters*. 2020 [cited 2020 April 2020]; Available from: http://www.kylesconverter.com/length/steps-to-meters.

58. Hamer, M. and Y. Chida, *Walking and primary prevention: a meta-analysis of prospective cohort studies.* Br J Sports Med, 2008. **42**(4): p. 238-43.

59. Hupin, D., et al., *Even a low-dose of moderate-to-vigorous physical activity reduces mortality by 22% in adults aged >/=60 years: a systematic review and meta-analysis.* Br J Sports Med, 2015. **49**(19): p. 1262-7.

60. Kelly, P., et al., *Systematic review and meta-analysis of reduction in all-cause mortality from walking and cycling and shape of dose response relationship.* Int J Behav Nutr Phys Act, 2014. **11**: p. 132.

61. Lollgen, H., A. Bockenhoff, and G. Knapp, *Physical activity and all-cause mortality: an updated meta-analysis with different intensity categories.* Int J Sports Med, 2009. **30**(3): p. 213-24.

62. Nocon, M., et al., *Association of physical activity with all-cause and cardiovascular mortality: a systematic review and meta-analysis.* Eur J Cardiovasc Prev Rehabil, 2008. **15**(3): p. 239-46.

63. Samitz, G., M. Egger, and M. Zwahlen, *Domains of physical activity and all-cause mortality: systematic review and dose-response meta-analysis of cohort studies.* Int J Epidemiol, 2011. **40**(5): p. 1382-400.

64. Woodcock, J., et al., *Non-vigorous physical activity and all-cause mortality: systematic review and meta-analysis of cohort studies.* Int J Epidemiol, 2011. **40**(1): p. 121-38.

65. Arem, H., et al., *Leisure Time Physical Activity and Mortality: A Detailed Pooled Analysis of the Dose-Response Relationship.* JAMA Internal Medicine, 2015. **175**(6): p. 959-967.

66. Holger Möller, F.H., Anurag Sharma and Lennert Veerman, *Proposal for a preferred method to cost the health benefits of active transport in New South Wales: Project Report*. 2019.

67. Price Water Cooperhouse, *Economic Valuation of Safety Benefits, Serious Injuries, Final Report, ,* R.a.T. Authority, Editor. 2005.

68. Australian Government, *Best Practice Regulation Guidance Note. Value of statistical life*, Department of the Prime Minister and Cabinet, Editor. 2019, Office of Best Practice Regulation.

69. Jonathan Karnon, H.H.A.A., Laura C Edney *New cancer drugs are very expensive - here's how we work out value for our money*. 2015.

70. Transport for New South Wales, *Principles and Guidelines for Economic Appraisal of Transport Investments and Initiatives. Transport Economic Appraisal Guidelines*. 2018, Evaluation and Assurance, Finance and Investment: https://www.transport.nsw.gov.au/projects/project-delivery-requirements/evaluation-and-assurance/resources.

71. Abelson P, *Establishing a Monetary Value for Lives Saved: Issues and Controversies. Working papers in cost-benefit analysis*. 2008, Office of Best Practice Regulation, Department of Finance and Deregulation.

72. Vos T, C.R., Barendregt J, Mihalopoulos C, Veerman JL, Magnus A, Cobiac L, Bertram MY, and A.P.T. Wallace AL, *Assessing Cost-Effectiveness in Prevention (ACE–Prevention): Final Report*. September 2010.

73. Roger Vickerman, R.B.N., Dick Ettema *Value of life and injuries*, in *Elsevier Science*, R. Vickerman, Editor. In Press, Elsevier Science: International Encyclopaedia of Transportation. p. 7,136 pages.

74. NSW Treasury, *A Note on the Discount Rate in the Context of Health.*, Jane Cheung, Editor. 2020.

75. Keeler, E.B. and S. Cretin, *Discounting of Life-Saving and Other Nonmonetary Effects.* Management Science, 1983. **29**(3): p. 300-306.

76. Brouwer, W.B.F., et al., *Need for differential discounting of costs and health effects in cost effectiveness analyses.* BMJ, 2005. **331**(7514): p. 446.

77. Government of UK, *The Green Book. Central Government Guidance on Appraisal and Evaluation*, HM Treasury, Editor. 2020: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/938046/The_Green_Book_2020.pdf.

78. Attema, A.E., W.B.F. Brouwer, and K. Claxton, *Discounting in Economic Evaluations.* PharmacoEconomics, 2018. **36**(7): p. 745-758.

79. Terrill M and Batrouney H, *Unfreezing discount rates. Transport infrastructure for tomorrow*. 2018, Grattan Institute: https://grattan.edu.au/report/unfreezing-discount-rates-transport-infrastructure-for-tomorrow/.

80. Pigou AC, *The economics of welfare*. 1920, London McMillan & Co.

81. Greg Jericho, *With borrowing effectively free, why be scared of government debt?*, in *The Guardian*. 8 March 2020: https://www.theguardian.com/business/commentisfree/2020/mar/08/with-borrowing-effectively-free-why-be-scared-of-government-debt.

82. Trading Economics. *Australia Interest Rate*. 2020 [cited 2020 November]; Available from: https://tradingeconomics.com/australia/interest-rate#:~:text=Looking%20forward%2C%20we%20estimate%20Interest,according%20to%20our%20econometric%20models.

83. Queensland Goverment. *Queensland Household Travel Survey summary reports*. 2016; Available from: https://www.tmr.qld.gov.au/Community-and-environment/Research-and-education/Queensland-Travel-Survey.

84. Barendregt, J.J. and J.L. Veerman, *Categorical versus continuous risk factors and the calculation of potential impact fractions.* Journal of Epidemiology and Community Health, 2010. **64**(3): p. 209.

85. Foley, L., et al., *Changes in active commuting and changes in physical activity in adults: a cohort study.* The international journal of behavioral nutrition and physical activity, 2015. **12**: p. 161-161.

86. Foley, L., et al., *Patterns of health behaviour associated with active travel: a compositional data analysis.* International Journal of Behavioral Nutrition and Physical Activity, 2018. **15**(1): p. 26.

87. Lachapelle, U., et al., *Active Transportation by Transit-Dependent and Choice Riders and Potential Displacement of Leisure Physical Activity.* Journal of Planning Education and Research, 2015. **36**(2): p. 225-238.

88. Henriques-Neto, D., et al., *Active Commuting and Physical Fitness: A Systematic Review.* International journal of environmental research and public health, 2020. **17**(8): p. 2721.

89. Infrastructure New South Wales, *Business Case Toolkit*. 2019: http://www.infrastructure.nsw.gov.au/project-assurance/resources/business-case-toolkit/.

90. NSW Government, *Transport for NSW Economic Parameter Values* Transport for New South Wales, Editor. 2019, Evaluation & Assurance Group Finance & Investment Corporate Services https://www.transport.nsw.gov.au/news-and-events/reports-and-publications/transport-for-nsw-economic-parameter-values.

91. Plowden, B., *Creating healthy streets for sustainable cities – delivering public health benefits through redesigning London's streets.* Cities & Health, 2019: p. 1-6.

92. Transport for New South Wales, *Principles and Guidelines: Economic Appraisal of Transport Investments and Initiatives. Transport Economic Appraisal Guidelines*. 2016, Evaluation and Benefits, Finance and Investment: https://www.transport.nsw.gov.au/principles-and-guidelines-economic-appraisal-of-transport-investments-and-initiatives.

93. Möller, H., et al., *What Is the Best Practice Method for Quantifying the Health and Economic Benefits of Active Transport?* International Journal of Environmental Research and Public Health, 2020. **17**(17): p. 6186.

94. Barendregt, J.J., et al., *A generic model for the assessment of disease epidemiology: the computational basis of DisMod II.* Population Health Metrics, 2003. **1**(1): p. 4-4.

95. Pluddemann, A., et al., *Redefining rapid reviews: a flexible framework for restricted systematic reviews.* BMJ Evid Based Med, 2018. **23**(6): p. 201-203.

96. David Moher, L.S., Mike Clarke, Davina Ghersi, Alessandro Liberati, Mark Petticrew, Paul Shekelle, Lesley A Stewart and PRISMA-P Group *Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement*. 2015.

97. Shea, B.J., et al., *AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both.* BMJ, 2017. **358**: p. j4008.

98. Kojima, N., et al., *Predictors of self-reported knee osteoarthritis in community-dwelling older women in Japan: A cross-sectional and longitudinal cohort study.* Archives of Gerontology and Geriatrics, 2017. **73**: p. 125-132.

99. Wang, Y., et al., *Is physical activity a risk factor for primary knee or hip replacement due to osteoarthritis? A prospective cohort study.* Journal of Rheumatology, 2011. **38**(2): p. 350-357.

100. Barry, V.W., et al., *Fitness vs. fatness on all-cause mortality: a meta-analysis.* Prog Cardiovasc Dis, 2014. **56**(4): p. 382-90.

101. Biswas, A., et al., *Sedentary time and its association with risk for disease incidence, mortality, and hospitalization in adults: a systematic review and meta-analysis.* Ann Intern Med, 2015. **162**(2): p. 123-32.

102. Chastin, S.F.M., et al., *How does light-intensity physical activity associate with adult cardiometabolic health and mortality? Systematic review with meta-analysis of experimental and observational studies.* Br J Sports Med, 2019. **53**(6): p. 370-376.

103. Chau, J.Y., et al., *Daily sitting time and all-cause mortality: a meta-analysis.* PLoS One, 2013. **8**(11): p. e80000.

104. Cooper, R., D. Kuh, and R. Hardy, *Objectively measured physical capability levels and mortality: systematic review and meta-analysis.* Bmj, 2010. **341**: p. c4467.

105. Cunningham, C., et al., *Consequences of physical inactivity in older adults: A systematic review of reviews and meta-analyses.* Scandinavian journal of medicine & science in sports, 2020.

106. Ekelund, U., et al., *Does physical activity attenuate, or even eliminate, the detrimental association of sitting time with mortality? A harmonised meta-analysis of data from more than 1 million men and women.* The Lancet, 2016. **388**(10051): p. 1302-1310.

107. Fogelholm, M., *Physical activity, fitness and fatness: relations to mortality, morbidity and disease risk factors. A systematic review.* Obes Rev, 2010. **11**(3): p. 202-21.

108. Karmisholt, K. and P.C. Gotzsche, *Physical activity for secondary prevention of disease. Systematic reviews of randomised clinical trials.* Dan Med Bull, 2005. **52**(2): p. 90-4.

109. Ku, P.W., et al., *Device-measured light-intensity physical activity and mortality: A meta-analysis.* Scandinavian journal of medicine & science in sports, 2020. **30**(1): p. 13-24.

110. Lacombe, J., et al., *The impact of physical activity and an additional behavioural risk factor on cardiovascular disease, cancer and all-cause mortality: a systematic review.* BMC Public Health, 2019. **19**(1): p. 900.

111. Liu, B., et al., *Usual walking speed and all-cause mortality risk in older people: A systematic review and meta-analysis.* Gait Posture, 2016. **44**: p. 172-7.

112. Loef, M. and H. Walach, *The combined effects of healthy lifestyle behaviors on all cause mortality: a systematic review and meta-analysis.* Prev Med, 2012. **55**(3): p. 163-70.

113. Patterson, R., et al., *Sedentary behaviour and risk of all-cause, cardiovascular and cancer mortality, and incident type 2 diabetes: a systematic review and dose response meta-analysis.* Eur J Epidemiol, 2018. **33**(9): p. 811-829.

114. Qiu, N.M. and Z.Q. Meng, *Relationship between physical exercise and sudden death and the physiological mechanism.* Chinese Journal of Clinical Rehabilitation, 2005. **9**(48): p. 160-161.

115. Rezende, L.F.M., et al., *All-Cause Mortality Attributable to Sitting Time: Analysis of 54 Countries Worldwide.* American Journal of Preventive Medicine, 2016. **51**(2): p. 253-263.

116. Warburton, D.E.R. and S.S.D. Bredin, *Health benefits of physical activity: a systematic review of current systematic reviews.* Curr Opin Cardiol, 2017. **32**(5): p. 541-556.

117. Yerrakalva, D., R. Mullis, and J. Mant, *The associations of "fatness," "fitness," and physical activity with all-cause mortality in older adults: A systematic review.* Obesity (Silver Spring), 2015. **23**(10): p. 1944-56.

118. Zhao, R., et al., *The Dose-Response Associations of Sedentary Time with Chronic Diseases and the Risk for All-Cause Mortality Affected by Different Health Status: A Systematic Review and Meta-Analysis.* Journal of Nutrition, Health and Aging, 2020. **24**(1): p. 63-70.

119. Ahn, S. and A.L. Fedewa, *A meta-analysis of the relationship between children's physical activity and mental health.* J Pediatr Psychol, 2011. **36**(4): p. 385-97.

120. Dale, L.P., et al., *Physical activity and depression, anxiety, and self-esteem in children and youth: An umbrella systematic review.* Mental Health and Physical Activity, 2019. **16**: p. 66-79.

121. Rebar, A.L., et al., *A meta-meta-analysis of the effect of physical activity on depression and anxiety in non-clinical adult populations.* Health Psychol Rev, 2015. **9**(3): p. 366-78.

122. Schuch, F.B., et al., *Are lower levels of cardiorespiratory fitness associated with incident depression? A systematic review of prospective cohort studies.* Prev Med, 2016. **93**: p. 159-165.

123. White, R.L., et al., *Domain-Specific Physical Activity and Mental Health: A Meta-analysis.* Am J Prev Med, 2017. **52**(5): p. 653-666.

124. Zimmermann, M., et al., *Modifiable risk and protective factors for anxiety disorders among adults: A systematic review.* Psychiatry Research, 2020. **285**.

125. Allen, K.D. and Y.M. Golightly, *State of the evidence.* Curr Opin Rheumatol, 2015. **27**(3): p. 276-83.

126. Alzahrani, H., et al., *The effectiveness of incidental physical activity interventions compared to other interventions in the management of people with low back pain: A systematic review and meta-analysis of randomised controlled trials.* Phys Ther Sport, 2019. **36**: p. 34-42.

127. Bean, J.F., A. Vora, and W.R. Frontera, *Benefits of exercise for community-dwelling older adults.* Arch Phys Med Rehabil, 2004. **85**(7 Suppl 3): p. S31-42; quiz S43-4.

128. Bennell, K., et al., *Exercise and osteoarthritis: cause and effects.* Compr Physiol, 2011. **1**(4): p. 1943-2008.

129. Booth, F.W., C.K. Roberts, and M.J. Laye, *Lack of exercise is a major cause of chronic diseases.* Compr Physiol, 2012. **2**(2): p. 1143-211.

130. Bosomworth, N.J., *Exercise and knee osteoarthritis: benefit or hazard?* Can Fam Physician, 2009. **55**(9): p. 871-8.

131. Burton, A.K., et al., *How to prevent low back pain.* Best Practice and Research: Clinical Rheumatology, 2005. **19**(4): p. 541-555.

132. Curl, W.W., *Aging and exercise: are they compatible in women?* Clin Orthop Relat Res, 2000(372): p. 151-8.

133. Dean, E. and A. Soderlund, *What is the role of lifestyle behaviour change associated with non-communicable disease risk in managing musculoskeletal health conditions with special reference to chronic pain?* BMC Musculoskelet Disord, 2015. **16**: p. 87.

134. Dugan, S.A., *Exercise for health and wellness at midlife and beyond: balancing benefits and risks.* Phys Med Rehabil Clin N Am, 2007. **18**(3): p. 555-75, xi.

135. Gardiner, B.S., et al., *Predicting Knee Osteoarthritis.* Ann Biomed Eng, 2016. **44**(1): p. 222-33.

136. Jones, G., M.G. Schultz, and D. Dore, *Physical activity and osteoarthritis of the knee: can MRI scans shed more light on this issue?* Phys Sportsmed, 2011. **39**(3): p. 55-61.

137. King, J., F.A. Reynolds, and L.H. De Souza, *Advancing knowledge of the causes of non-specific low back pain for primary prevention: A systematic review.* Physiotherapy (United Kingdom), 2011. **97**: p. eS613.

138. Lefevre-Colau, M.M., et al., *Is physical activity, practiced as recommended for health benefit, a risk factor for osteoarthritis?* Ann Phys Rehabil Med, 2016. **59**(3): p. 196-206.

139. Øiestad, B.E., et al., *Risk factors for episodes of back pain in emerging adults. A systematic review.* European Journal of Pain (United Kingdom), 2020. **24**(1): p. 19-38.

140. Semanik, P.A., R.W. Chang, and D.D. Dunlop, *Aerobic activity in prevention and symptom control of osteoarthritis.* Pm r, 2012. **4**(5 Suppl): p. S37-44.

141. Stevens-Lapsley, J.E. and W.M. Kohrt, *Osteoarthritis in women: effects of estrogen, obesity and physical activity.* Womens Health (Lond), 2010. **6**(4): p. 601-15.

142. Urquhart, D.M., et al., *Factors that may mediate the relationship between physical activity and the risk for developing knee osteoarthritis.* Arthritis Res Ther, 2008. **10**(1): p. 203.

143. Urquhart, D.M., et al., *What is the effect of physical activity on the knee joint? A systematic review.* Med Sci Sports Exerc, 2011. **43**(3): p. 432-42.

144. Vignon, E., et al., *Osteoarthritis of the knee and hip and activity: a systematic international review and synthesis (OASIS).* Joint Bone Spine, 2006. **73**(4): p. 442-55.

145. Vuori, I.M., *Dose-response of physical activity and low back pain, osteoarthritis, and osteoporosis.* Med Sci Sports Exerc, 2001. **33**(6 Suppl): p. S551-86; discussion 609-10.

146. Øverås, C.K., et al., *Association between objectively measured physical behaviour and neck- and/or low back pain: A systematic review.* European Journal of Pain (United Kingdom), 2020.

147. Kraus, W.E., et al., *Physical Activity, All-Cause and Cardiovascular Mortality, and Cardiovascular Disease.* Med Sci Sports Exerc, 2019. **51**(6): p. 1270-1281.

148. Transport., A.G.D.o.I.a., *Walking, Riding and Access to Public Transport; Supporting Active Travel in Australian Communities* 2013, Commonwealth of Australia.

149. Di Blasio, A., et al., *Walking training in postmenopause: effects on both spontaneous physical activity and training-induced body adaptations.* Menopause, 2012. **19**(1): p. 23-32.

150. Longo, A., et al., *Demand response to improved walking infrastructure: A study into the economics of walking and health behaviour change.* Soc Sci Med, 2015. **143**: p. 107-16.

151. Foley, L., et al., *Changes in active commuting and changes in physical activity in adults: a cohort study.* Int J Behav Nutr Phys Act, 2015. **12**: p. 161.

152. Gomersall, S.R., et al., *In search of lost time: When people undertake a new exercise program, where does the time come from? A randomized controlled trial.* J Sci Med Sport, 2015. **18**(1): p. 43-8.

153. Goodman, A., et al., *New walking and cycling routes and increased physical activity: one- and 2-year findings from the UK iConnect Study.* American journal of public health, 2014. **104**(9): p. e38-46.

154. Salvo, G., et al., *A Mixed Methods Study on the Barriers and Facilitators of Physical Activity Associated with Residential Relocation.* J Environ Public Health, 2018. **2018**: p. 1094812.

155. Thielman, J., et al., *Neighborhood walkability: Differential associations with self-reported transport walking and leisure-time physical activity in Canadian towns and cities of all sizes.* Prev Med, 2015. **77**: p. 174-80.

156. Sahlqvist, S., et al., *Change in active travel and changes in recreational and total physical activity in adults: longitudinal findings from the iConnect study.* The international journal of behavioral nutrition and physical activity, 2013. **10**: p. 28-28.

157. van Tienoven, T.P., et al., *Active work, passive leisure? Associations between occupational and non-occupational physical activity on weekdays.* Social Science Research, 2018. **76**: p. 1-11.

158. Tigbe, W.W., M.E.J. Lean, and M.H. Granat, *A physically active occupation does not result in compensatory inactivity during out-of-work hours.* Preventive Medicine, 2011. **53**(1): p. 48-52.

159. Clemes, S.A., et al., *Sitting time and step counts in office workers.* Occupational Medicine, 2014. **64**(3): p. 188-192.

160. Rowland, T.W., *The biological basis of physical activity.* Medicine & Science in Sports & Exercise, 1998. **30**(3).

161. Gomersall, S.R., et al., *The ActivityStat Hypothesis.* Sports Medicine, 2013. **43**(2): p. 135-149.

162. Gomersall, S.R., et al., *Testing the activitystat hypothesis: a randomised controlled trial.* BMC public health, 2016. **16**(1): p. 900-900.

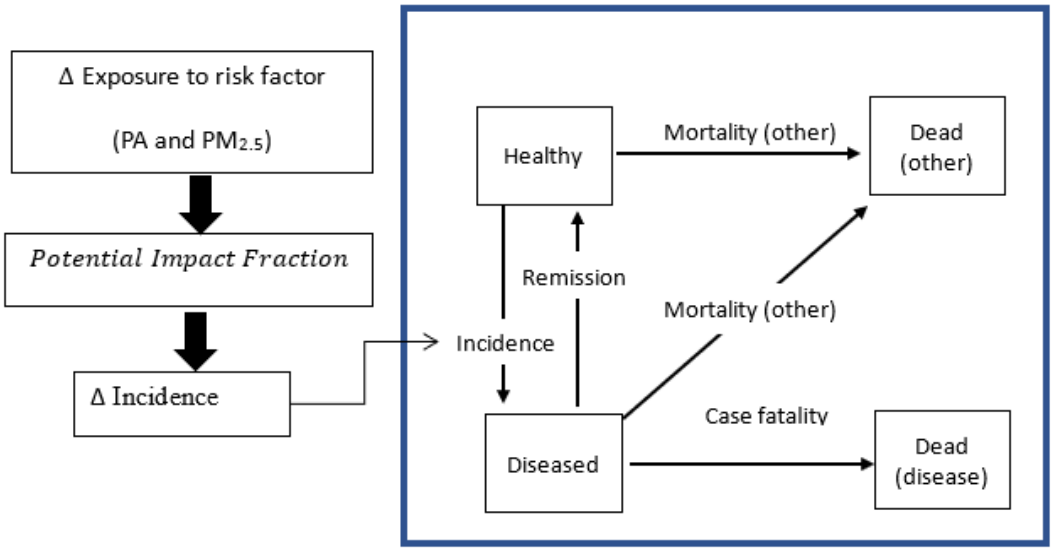
11. Appendices

Appendix A: Criteria identified in stakeholder consultation

Box 1 Desirable characteristics of model to cost the health benefits of active transport from stakeholder consultation

- Uses robust statistical model
- Allows for different types of active transport characterised by duration and intensity (energy expenditure)
- Allows for modelling of subgroups
- Models impacts at fine grained level (person / age group / gender) as opposed to whole of population
- Considers all exposures with sufficiently strong epidemiological evidence of causal relationship to health outcomes
- Considers all health outcomes with sufficiently strong epidemiological evidence and causal relationship to exposure
- Considers morbidity and mortality
- Models the impact of active transport over the life course, allowing for co-morbidities
- Produces outcome measures that can be used in cost benefit analysis
- Aligns with NSW Treasury Guidelines for Cost Benefit Analysis
- Considers health related economic outcomes such as health care costs, productivity costs, costs to others / society (in addition to health outcomes)

The fourteen included diseases were modelled applying a set of differential equations to describe the transition between four states (healthy, diseased, dead from the disease and dead from all other causes) [94] (Appendix Figure B-1). Transition probabilities among the four states reflect rates of incidence, remission, case fatality and background mortality. A change in exposure to the risk factors of interest (PA and $PM_{2.5}$) modifies incidence via the potential impact fraction calculation (PIF).



Appendix Figure B-1 Conceptual disease model

The model was used for each of the physical activity and $PM_{2.5}$ related diseases. The disease conceptual model, applied to each disease separately, has four health states (healthy, diseased, dead from the disease and dead from other causes) and transition hazards between health states [94]. The diseases conceptual model is that of a multi-state life table model (box in figure).

Double counting of health effects and impact on healthcare costs is avoided by the life table structure, and by ignoring disease-specific mortality when an impact of PA on ACM is applied.

When two or more health outcomes occur together, the model deals with the impact on healthcare costs correctly and any degree of double counting in the disease burden would be negligible. In the lifetable structure, at any age and sex, proportions of the population suffer from any of the conditions. For example, among 65-year-old males, 14% have diabetes and 9% have ischaemic heart disease. Health care costs are associated with the rate of new cases for cancers, and with the proportion of existing cases (prevalence) for all other conditions. This is consistent with the way the health care costs were apportioned in the original study (from AIHW): a total 'envelope' of costs was divided up over all health conditions while avoiding double-counting. Implicitly, any 'savings' from combining conditions was already accounted for; this would have lowered to per-case costs for the conditions

involved. Here, our model is consistent with the data input. All health care costs that are not related to the diseases that are explicitly modelled are divided by population numbers to come to an average annual cost for each age and sex (e.g., for 65-year-old males, this is $7,455 per person per year). For quality of life, we use 'disability weights' from the Global Burden of Disease. These are conceptualised as proportions of quality of life lost to a condition. Adding up these losses could in theory lead to a quality of life below 0, i.e., worse than death. However, that is in individuals, in which diseases cluster. In our model that does not occur as it has the population as unit, with a percentage for each modelled condition (prevalence). An intervention results in small shifts in these percentages with disease take place. Because these changes at the population level are so marginal for realistic interventions (and for our reference case scenario), the overlap of diseases would be marginal.

As mentioned above, double counting of health effects is also avoided by ignoring disease-specific mortality when an impact of PA on ACM is applied. Incorporating that direct link between physical activity and the risk of death accounts for the effects via the modelled diseases, but also includes any effects via other conditions that are in themselves to rare to result in significant findings in cohort studies, and the fact that while our model only takes into account an effect of physical activity on the risk for new cases of disease but ignores any effect on the mortality of persons already with that disease. For example, our model gives no mortality benefit for people with diabetes if they become more active. Mortality is also an outcome that can be reliable established and traced in cohort studies, as even for people who drop out of a study by not responding to the researchers, their death can often be recorded via mortality registries. Of course, in these observational studies, the estimation of the 'independent' effect of physical activity on this outcome depends on accurate measurement and statistical removal (adjustment) of competing causes of that outcome – but that is true for any outcome, and in our report we discuss the evidence for (and against) a causal interpretation of the association found in the pooled evidence.

Appendix C: Model input parameters, updates and data sources

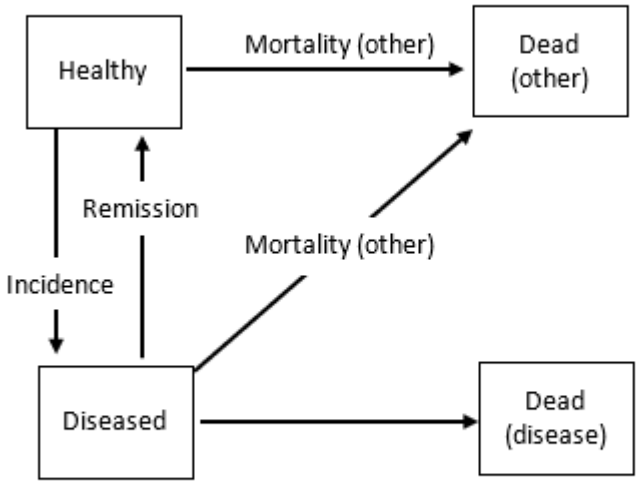
Data	**Data source original study**	**The data source for update**
Mortality rates and population numbers	Deaths, Australia, 2015 [Internet]. Australia Bureau of Statistics. 2016 http://www.abs.gov.au/ausstats/abs@.nsf/mf/3302.0. Estimated Resident Population By Single Year of Age Australia [Internet]. 2016 http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3101.0Dec%202015?OpenDocument	ABS 3302.0 - Deaths, Australia, 2017 https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3302.02017?OpenDocument 3101.0 - Australian Demographic Statistics, Dec 2018 https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3101.0Dec%202018?OpenDocument
Years lived with disability (YLD) (all causes and road trauma)	Institute of Health Metrics and Evaluation GBD Compare [Internet]. IHME, University of Washington. 2015 [cited 15 July 2016]. https://vizhub.healthdata.org/gbd-compare/.	Institute of Health Metrics and Evaluation, 2017 data, https://vizhub.healthdata.org/gbd-compare/
Incidence and case fatality modeled diseases	DisMod II Barendregt JJ. EpiGear International 2012 [cited 2015 1 Mar].	Update GBD data to 2017, update AIHW Cancer data in Australia to 2018, https://www.aihw.gov.au/reports

	http://www.epigear.com/index_files/prevent.html. Global Burden of Disease (GBD) 2013 data Global Burden of Disease (GBD) [Internet]. 2015 http://www.healthdata.org/gbd. Australian Institute of Health and Welfare (AIHW) data Australian Cancer Incidence and Mortality (ACIM) books [Internet]. 2016 http://www.aihw.gov.au/acim-books.	/cancer/cancer-data-in-australia/acim-books
Disability weights modeled diseases	Prevalence and years lived with disability from GBD 2013, Institute for Health Metrics and Evaluation (IHME). (2015). GBD Compare.	https://vizhub.healthdata.org/gbd-compare/
Relative risk, PA	Danaei G, Ding EL, Mozaffarian D, Taylor B, Rehm J, Murray CJL, et al. The preventable causes of death in the United States: comparative risk assessment of dietary, lifestyle, and metabolic risk factors. PLoS Med. 2009; 6(4):e1000058. https://doi.org/10.1371/journal.pmed.1000058 PMID: 19399161	Kyu HH, Bachman VF, Alexander LT, Mumford JE, Afshin A, Estep K, Veerman JL, Delwiche K, Iannarone ML, Moyer ML, Cercy K. Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: a systematic review and dose-response meta-analysis for the Global Burden of Disease Study 2013. BMJ 2016 Aug 9;354:i3857 Schuch FB, Vancampfort D, Firth J, et al. Physical Activity and Incident Depression: A Meta-Analysis of Prospective Cohort Studies. The American journal of psychiatry 2018; 175(7): 631-48. Kelly P, Kahlmeier S, Gotschi T, et al. Systematic review and meta-analysis of reduction in all-cause mortality from walking and cycling and shape of dose response relationship. The international journal of behavioural nutrition and physical activity 2014; 11: 132
Relative risks, ischaemic heart disease and ischaemic stroke due to diabetes	Asia Pacific Cohort Studies Collaboration. The Effects of Diabetes on the Risks of Major Cardiovascular Diseases and Death in the Asia-Pacific Region. Diabetes Care. 2003; 26(2):360±6. https://doi.org/10.2337/diacare.26.2.360 PMID: 12547863	

Relative risk, PM2.5	World Health Organization. WHO Expert Meeting: Methods and tools for assessing the health risks of air pollution at local, national and international level. Copenhagen: WHO Regional Office for Europe, 2014 Hamra GB, Guha N, Cohen A, Laden F, Raaschou-Nielsen O, Samet J, et al. Outdoor particulate matter exposure and lung cancer: a systematic review and meta-analysis. Environ Health Perspect. 2014; 122:906±11. https://doi.org/10.1289/ehp.1408092 PMID: 24911630	Asia Pacific Cohort Studies Collaboration. The Effects of Diabetes on the Risks of Major Cardiovascular Diseases and Death in the Asia-Pacific Region. Diabetes Care. 2003; 26(2):360±6. https://doi.org/10.2337/diacare. 26.2.360 PMID: 12547863
Mediating effect of diabetes in the causal pathway between physical and ischemic heart disease and ischemic stroke	GBD 2013 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990±2013: A systematic analysis for the Global Burden of Disease Study 2013. The Lancet. 2015; 386(10010):2287±323. https://doi.org/10.1016/S0140-6736(15)00128-2 PMID: 26364544 page 711	Latest update: GBD 2017 Risk Factor Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet (London, England) 2018; 392(10159): 1923-94.
Physical activity categories	National Nutrition and Physical Activity Survey Basic Confidentialised Unit Record File (CURF) Australian Health Survey: Physical Activity, 2011±12 [Internet]. 2015 [http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/D4495467B7F7EB01CA257BAC0015F593?opendocument.	ABS 4364.0.55.001 - National Health Survey: First Results, 2017-18 https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/4364.0.55.001~2017-18~Main%20Features~Physical%20activity~115
MET-minutes (walking = 3.5, cycling = 6.8, moderate PA = 5, vigorous PA = 7.5)	Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett DR Jr., Tudor-Locke C, et al. 2011 compendium of physical activities: a second update of codes and MET values. Med Sci Sports Exerc. 2011; 43(8):1575±81. Epub 2011/06/18. https://doi.org/10.1249/MSS.0b013e31821ece12 PMID: 21681120 for walking and cycling; Australian Health Survey: Physical Activity, 2011±12 [Internet]. 2015 http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/D4495467B7F7EB01CA257BAC0015F593? opendocument. for walking, moderate and vigorous PA	ABS 4364.0.55.001 - National Health Survey: First Results, 2017-18 https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/4364.0.55.001~2017-18~Main%20Features~Physical%20activity~115
Health care costs	Australian Institute of Health and Welfare. Disease costs and impact study	AIHW Disease expenditure in Australia

	data. Australian Institute of Health and Welfare, 2001. All diseases except COPD were indexed to 2013 using AIHW-reported health sector indices. Denominators for calculating per case costs (incidence, prevalence and years lived with disability) Global Burden of Disease Study 2015 (GBD 2015) Results [Internet]. 2016 [cited 19 October 2016]. http://ghdx.healthdata.org/gbd-results-tool.	https://www.aihw.gov.au/reports/health-welfare-expenditure/disease-expenditure-australia/data Institute of Health Metrics and Evaluation, 2017 data, https://vizhub.healthdata.org/gbd
Discount rate for health care costs	Murray CJL, Ezzati M, Flaxman AD, Lim S, Lozano R, Michaud C, et al. GBD 2010: design, definitions, and metrics. The Lancet. 2012; 380(9859):2063±6. https://doi.org/10.1016/S0140-6736(12)61899-6 for health; Gold MR. Cost-effectiveness in health and medicine. New York: Oxford University Press; 1996.et al. for health care costs	NSW Government Treasury. Policy and Guidelines Paper NSW Government- Guide to Cost-Benefit Analysis: NSW Government - The Treasury, 2017.

The model requires input parameters: incidence, prevalence, remission and case fatality rates for each disease. Remission and case fatality rates are not provided by the GBD study and were derived using DISMOD II. The conceptual model of DisMod II is that of a MSLT (Appendix Figure D-1). "*Healthy people, defined as people unaffected by the disease being modelled, are subject to an incidence hazard, and may become diseased. When diseased they are subject to a hazard of dying from the disease, the case fatality, and to a hazard of recovery from the disease, called remission. Both healthy and diseased people are subject to the same mortality hazard from all other causes*" [94, p. 2].



Appendix Figure D-1 Conceptual disease model in the MSLT

DisMod II generates age and sex specific internally consistent estimates for the disease modelled. At least three of the following input parameters are required for DisMod II to estimate model input values: incidence, prevalence, remission, case fatality, duration, mortality and relative risks on mortality (for the modelled disease). We used the input parameters; incidence, prevalence and mortality. Data for the three input parameters were obtained from the GBD 2017 study (Appendix table D-1) [17]. We have provided a summary of all data used for DisMod II inputs and where additional procedures were applied to generate the data, a description of the procedures used has been provided below in Appendix table D-1.

Appendix table D-1 DisMod II input data and procedures for modelled diseases

Disease	Collection input parameters [a]	Dataset input parameters [a]	Procedures [b]
Breast cancer	Population Mortality rates	Incidence, Prevalence, Mortality	No additional procedure applied
Colon and rectum cancer	Population Mortality rates	Incidence, Prevalence, Mortality	No additional procedure applied
Tracheal, bronchus, and lung cancer	Population Mortality rates	Incidence, Prevalence, Mortality	Remission set to zero Weights: halfway bar weight for prevalence and mortality input parameters
Diabetes mellitus type 2	Population Mortality rates	Incidence, Prevalence, Mortality	No additional procedure applied
Ischemic heart disease	Population Mortality rates	Incidence, Prevalence, Remission, Mortality	Remission set to zero Weights: ignore incidence
Ischemic stroke	Population Mortality rates	Incidence, Prevalence, Mortality	No additional procedure applied
Intracerebral haemorrhage	Population Mortality rates	Incidence, Prevalence, Mortality	No additional procedure applied
Subarachnoid haemorrhage	Population Mortality rates	Incidence, Prevalence, Mortality	No additional procedure applied
Chronic obstructive pulmonary disease	Population Mortality rates	Incidence, Prevalence, Remission, Mortality	Remission set to zero Highest weight to prevalence and mortality. Lowest weight to incidence.
Lower respiratory infections	Population Mortality rates	Incidence, Prevalence, Mortality	Constraints applied: Incidence- 5, Remission- 500, Case fatality- 100
Anxiety disorders	Population Mortality rates	Incidence, Prevalence, Case fatality, Mortality	Weights: Case fatality set to exact zero Constraints applied: hazards maximum, Case fatality- value=0, enforced=0
Depressive disorders	Population Mortality rates	Incidence, Prevalence, Case fatality, Mortality	Constraints applied: Incidence- 5, Remission- 500, Case fatality- 0, Weights: Case fatality set to exact zero
Osteoarthritis	Population Mortality rates	Incidence, Prevalence, Case fatality, Mortality	Highest weighting for prevalence

			Constraints applied: Incidence-5, Remission- 500, Case fatality- 0 Weights: case fatality -Exact
Low back pain	Population Mortality rates	Incidence, Prevalence, Case fatality, Mortality	Highest weighting for prevalence Constraints applied: Incidence-5, Remission- 500, Case fatality- 0 Weights: case fatality -Exact

[a] Dataset input parameters sourced from the GBD Results tool available from the Institute for Health Metrics and Evaluation (IHME), 2018 webpage [17]. [b] These procedures are based on options given by DisMod II.

We derived disability weights (DW) from disease specific years lived with disability (YLDs) and disease specific prevalence by five-year age group and sex. Data for YLDs and disease prevalence were obtained from the online GBD Results tool [17].

Disability weights were derived by mean YLD for each age/sex group divided by the mean prevalence of that disease. For this study we had data at the cause level (e.g. IHD) instead of sequela level (e.g. myocardial infarction (MI), chronic IHD, angina, asymptomatic IHD following MI, acute MI). Our calculations are based on the GBD methods for estimating YLDs as the disease prevalence multiplied by disability weights [16].

We then adjusted for disability due to other diseases in each age group. We inflated the result to account for the fact that due to the presence of other health conditions, the starting point for the calculation of the YLDs was less than full health. So, we inflated by '1-pYLD / pop', whereby pYLD stands for the total YLD for all diseases except the one under scrutiny, and pop is the population number.

Adjusted DW = DW / (1- (pYLD /Pop))

Appendix E: Annual disease cost per case

An outline of the annual disease cost

Calculated health care cost	Outcomes / Description
Cost per incident case of disease	Lower respiratory tract infections, Breast cancer, Colon rectum cancer, Lung cancer
Cost per prevalent case of disease	Coronary heart disease, Type 2 diabetes, Chronic obstructive pulmonary disease, Ischemic stroke, Intracerebral haemorrhage, Subarachnoid haemorrhage, Depressive disorders, Anxiety disorders, Osteoarthritis, Low back pain
Costs per prevalent YLD	Road traffic injury - Pedestrians, Road traffic injury - pedal cyclists, Road traffic injury -Motorcyclist, Road traffic injury - Motor vehicle occupants
Costs for all other diseases in added life years	Overall health care costs minus costs of diseases and injuries included in the model

Annual Disease cost per case

Sex	Age (years)	Coronary heart disease (IHD-in model) Cost per prevalent case of disease	Type 2 diabetes Cost per prevalent case of disease	Chronic obstructive pulmonary disease Cost per prevalent case of disease	Stroke (in model= Ischemic stroke, Intracerebral haemorrhage, Subarachnoid haemorrhage,) Cost per prevalent case of disease	Lower respiratory infections Cost per incident case of disease	Depressive disorders Cost per prevalent case of disease
Male	Under 5	$ -	$ -	$ 48,032	$ 38,253	$ 2,970	$ -
Male	5-9	$ -	$ -	$ 11,200	$ 6,632	$ 857	$ 19,077
Male	10-14	$ -	$ 277,589	$ 6,264	$ 3,315	$ 1,057	$ 2,811
Male	15-19	$ 47,283	$ 38,391	$ 4,003	$ 4,469	$ 1,358	$ 1,636
Male	20-24	$ 16,323	$ 4,802	$ 2,306	$ 2,813	$ 1,363	$ 1,368
Male	25-29	$ 10,941	$ 1,885	$ 1,704	$ 2,631	$ 1,357	$ 1,381
Male	30-34	$ 7,832	$ 1,242	$ 1,320	$ 2,499	$ 1,563	$ 1,549
Male	35-39	$ 6,409	$ 1,001	$ 1,012	$ 2,832	$ 1,882	$ 1,639
Male	40-44	$ 5,993	$ 1,052	$ 855	$ 2,508	$ 2,407	$ 1,804
Male	45-49	$ 4,761	$ 917	$ 715	$ 2,415	$ 2,685	$ 1,885
Male	50-54	$ 4,871	$ 970	$ 698	$ 2,412	$ 3,561	$ 2,062
Male	55-59	$ 4,210	$ 946	$ 695	$ 2,326	$ 4,542	$ 2,245
Male	60-64	$ 4,133	$ 992	$ 714	$ 2,344	$ 5,180	$ 2,377
Male	65-69	$ 3,902	$ 1,067	$ 774	$ 2,541	$ 5,303	$ 2,479
Male	70-74	$ 3,434	$ 1,012	$ 819	$ 2,354	$ 5,638	$ 2,757
Male	75-79	$ 3,454	$ 1,170	$ 936	$ 2,725	$ 7,468	$ 3,218
Male	80-84	$ 3,115	$ 1,189	$ 1,056	$ 2,748	$ 6,258	$ 3,550
Male	85 and over	$ 2,620	$ 1,113	$ 1,060	$ 2,381	$ 5,317	$ 3,276
Female	Under 5	$ -	$ -	$ 33,204	$ 15,865	$ 2,583	$ -
Female	5-9	$ -	$ -	$ 8,640	$ 3,167	$ 865	$ 9,962
Female	10-14	$ -	$ 403,496	$ 4,700	$ 2,566	$ 927	$ 1,929
Female	15-19	$ 68,856	$ 29,801	$ 4,165	$ 2,726	$ 1,155	$ 1,798

Female	20-24	$ 26,512	$ 5,137	$ 3,116	$ 2,443	$ 1,287	$ 1,312
Female	25-29	$ 17,283	$ 2,894	$ 2,337	$ 2,371	$ 1,560	$ 1,422
Female	30-34	$ 12,232	$ 2,088	$ 1,795	$ 2,638	$ 1,916	$ 1,680
Female	35-39	$ 7,745	$ 1,416	$ 1,221	$ 2,175	$ 2,185	$ 1,743
Female	40-44	$ 6,160	$ 1,111	$ 936	$ 2,545	$ 2,466	$ 1,855
Female	45-49	$ 4,739	$ 917	$ 735	$ 2,013	$ 2,550	$ 1,983
Female	50-54	$ 4,604	$ 942	$ 754	$ 1,935	$ 3,538	$ 2,111
Female	55-59	$ 3,712	$ 876	$ 750	$ 1,770	$ 4,372	$ 2,185
Female	60-64	$ 3,624	$ 855	$ 755	$ 1,707	$ 4,886	$ 2,212
Female	65-69	$ 3,460	$ 874	$ 816	$ 1,693	$ 5,042	$ 2,298
Female	70-74	$ 3,173	$ 806	$ 822	$ 1,726	$ 5,191	$ 2,526
Female	75-79	$ 3,348	$ 918	$ 885	$ 2,131	$ 6,878	$ 3,334
Female	80-84	$ 3,271	$ 926	$ 894	$ 2,332	$ 5,773	$ 3,748
Female	85 and over	$ 2,731	$ 842	$ 815	$ 2,173	$ 4,507	$ 3,034

Sex	Age (years)	Anxiety disorders Cost per prevalent case of disease	Osteoarthritis Cost per prevalent case of disease	Low back pain Cost per prevalent case of disease	Breast cancer Cost per incident case of disease	Bowel cancer (Colon rectum cancer) Cost per incident case of disease	Lung cancer Cost per incident case of disease
Male	Under 5	$ 16,424	$ -	$ -	$ -	$ -	$ -
Male	5-9	$ 2,132	$ -	$ 3,843	$ -	$ -	$ -
Male	10-14	$ 961	$ -	$ 848	$ -	$ -	$ -
Male	15-19	$ 1,005	$ -	$ 478	$ 40,388	$ 471,596	$ 2,533,193
Male	20-24	$ 993	$ -	$ 389	$ 83,334	$ 123,828	$ 550,301
Male	25-29	$ 1,042	$ -	$ 462	$ 28,644	$ 104,046	$ 225,015
Male	30-34	$ 1,135	$ 2,117	$ 560	$ 80,601	$ 76,251	$ 140,109
Male	35-39	$ 1,235	$ 1,583	$ 612	$ 34,110	$ 51,116	$ 78,506
Male	40-44	$ 1,409	$ 1,631	$ 770	$ 83,332	$ 56,325	$ 65,410
Male	45-49	$ 1,317	$ 1,455	$ 767	$ 75,230	$ 47,898	$ 42,221
Male	50-54	$ 1,431	$ 1,819	$ 829	$ 83,455	$ 47,635	$ 35,652
Male	55-59	$ 1,497	$ 2,147	$ 911	$ 90,828	$ 44,354	$ 32,727
Male	60-64	$ 1,766	$ 2,627	$ 1,095	$ 69,924	$ 42,278	$ 33,867
Male	65-69	$ 2,307	$ 2,907	$ 1,362	$ 125,360	$ 41,161	$ 29,834
Male	70-74	$ 2,140	$ 2,643	$ 1,384	$ 84,782	$ 35,694	$ 26,427
Male	75-79	$ 2,648	$ 2,695	$ 1,527	$ 91,246	$ 35,782	$ 24,868
Male	80-84	$ 3,120	$ 2,110	$ 1,486	$ 117,840	$ 32,265	$ 20,742
Male	85 and over	$ 3,901	$ 1,231	$ 1,284	$ 40,953	$ 17,951	$ 15,894
Female	Under 5	$ 9,269	$ -	$ -	$ -	$ -	$ -
Female	5-9	$ 1,089	$ -	$ 2,929	$ -	$ -	$ -
Female	10-14	$ 838	$ -	$ 1,105	$ -	$ -	$ -
Female	15-19	$ 1,348	$ -	$ 573	$ 1,480,195	$ 510,398	$ 1,296,337
Female	20-24	$ 1,004	$ -	$ 410	$ 305,061	$ 181,627	$ 523,028
Female	25-29	$ 943	$ -	$ 486	$ 83,217	$ 136,195	$ 233,564
Female	30-34	$ 1,020	$ 3,562	$ 567	$ 66,113	$ 88,348	$ 183,229

Female	35-39	$	1,039	$	1,998	$	604	$	62,457	$	70,841	$	127,715
Female	40-44	$	1,101	$	1,395	$	695	$	70,392	$	78,304	$	108,450
Female	45-49	$	1,054	$	1,120	$	677	$	55,017	$	58,775	$	62,459
Female	50-54	$	1,170	$	1,404	$	790	$	58,462	$	57,300	$	53,606
Female	55-59	$	1,223	$	1,595	$	826	$	54,876	$	49,062	$	46,782
Female	60-64	$	1,317	$	1,961	$	1,001	$	52,402	$	47,471	$	39,882
Female	65-69	$	1,506	$	2,312	$	1,259	$	52,108	$	42,629	$	37,304
Female	70-74	$	1,586	$	2,192	$	1,441	$	47,763	$	35,006	$	31,277
Female	75-79	$	2,107	$	2,319	$	1,668	$	42,884	$	30,201	$	31,020
Female	80-84	$	2,560	$	1,888	$	1,487	$	37,824	$	27,863	$	25,681
Female	85 and over	$	2,982	$	1,068	$	1,423	$	17,112	$	16,253	$	17,505

Sex	Age (years)	Road traffic injury - Pedestrians Costs/ prevalent YLD	Road traffic injury - pedal cyclists Costs/prevalent YLD	Road traffic injury - Motorcyclist Costs/prevalent YLD	Road traffic injury - Motor vehicle occupants Costs/prevalent YLD
Male	Under 5	$ 141,250	$ 338,508	$ 246,309	$ 170,990
Male	5-9	$ 27,536	$ 144,311	$ 167,412	$ 48,034
Male	10-14	$ 17,282	$ 36,382	$ 258,072	$ 113,554
Male	15-19	$ 14,486	$ 27,829	$ 127,228	$ 70,481
Male	20-24	$ 10,802	$ 32,517	$ 67,821	$ 38,117
Male	25-29	$ 7,752	$ 22,070	$ 40,845	$ 25,194
Male	30-34	$ 5,893	$ 16,123	$ 27,162	$ 18,488
Male	35-39	$ 5,345	$ 13,142	$ 22,110	$ 15,796
Male	40-44	$ 4,445	$ 14,322	$ 17,408	$ 12,569
Male	45-49	$ 3,950	$ 11,988	$ 14,667	$ 10,893
Male	50-54	$ 4,268	$ 12,102	$ 14,385	$ 11,472
Male	55-59	$ 4,321	$ 11,803	$ 14,152	$ 11,780
Male	60-64	$ 4,653	$ 4,059	$ 2,598	$ 9,696
Male	65-69	$ 5,183	$ 4,656	$ 2,894	$ 10,996
Male	70-74	$ 5,548	$ 5,182	$ 3,193	$ 11,997
Male	75-79	$ 7,190	$ 7,497	$ 4,457	$ 16,102
Male	80-84	$ 7,292	$ 1,059	$ 669	$ 11,933
Male	85 and over	$ 9,481	$ 1,673	$ 992	$ 15,738
Female	Under 5	$ 88,608	$ 216,165	$ 82,101	$ 197,433
Female	5-9	$ 26,130	$ 56,431	$ 48,356	$ 56,147
Female	10-14	$ 22,771	$ 22,748	$ 79,414	$ 114,395
Female	15-19	$ 14,342	$ 14,612	$ 39,460	$ 67,432
Female	20-24	$ 6,261	$ 17,503	$ 22,135	$ 32,322
Female	25-29	$ 4,980	$ 13,641	$ 16,584	$ 24,329
Female	30-34	$ 4,388	$ 11,699	$ 13,529	$ 20,369
Female	35-39	$ 4,050	$ 9,849	$ 11,399	$ 17,696
Female	40-44	$ 3,209	$ 6,528	$ 5,649	$ 11,572
Female	45-49	$ 2,916	$ 5,771	$ 4,970	$ 10,195
Female	50-54	$ 3,403	$ 6,638	$ 5,455	$ 11,107
Female	55-59	$ 3,423	$ 6,714	$ 5,600	$ 11,057
Female	60-64	$ 3,447	$ 1,396	$ 784	$ 7,071
Female	65-69	$ 3,971	$ 1,672	$ 955	$ 8,295
Female	70-74	$ 4,411	$ 1,928	$ 1,083	$ 9,419
Female	75-79	$ 5,872	$ 2,873	$ 1,691	$ 12,885
Female	80-84	$ 2,432	$ 169	$ 568	$ 6,485
Female	85 and over	$ 3,660	$ 264	$ 813	$ 10,017

N.B. Costs are in Australian dollars, from the Disease Expenditure in Australia 2015-16 report prepared by the Australian Institute of Health and Welfare [21] . RTI = road traffic injury

Appendix F: Costs for all other diseases in the added life years

Sex	Age	(Overall health care costs minus costs of diseases and injuries included in the model) All other costs per person	
Male	Under 5	$	3,347
Male	5-9 years	$	1,237
Male	10-14 years	$	1,163
Male	15-19 years	$	1,504
Male	20-24 years	$	1,594
Male	25-29 years	$	1,677
Male	30-34 years	$	1,885
Male	35-39 years	$	2,156
Male	40-44 years	$	2,669
Male	45-49 years	$	2,895
Male	50-54 years	$	3,812
Male	55-59 years	$	4,612
Male	60-64 years	$	5,824
Male	65-69 years	$	7,471
Male	70-74 years	$	8,589
Male	75-79 years	$	11,494
Male	80-84 years	$	13,713
Male	85 years and over	$	15,611
Female	Under 5 years	$	2,792
Female	5-9 years	$	1,023
Female	10-14 years	$	1,063
Female	15-19 years	$	1,966
Female	20-24 years	$	2,700
Female	25-29 years	$	3,572
Female	30-34 years	$	4,470
Female	35-39 years	$	4,043
Female	40-44 years	$	3,619
Female	45-49 years	$	3,273
Female	50-54 years	$	3,943
Female	55-59 years	$	4,263
Female	60-64 years	$	4,992
Female	65-69 years	$	6,260
Female	70-74 years	$	7,190
Female	75-79 years	$	9,564
Female	80-84 years	$	11,397
Female	85 years and over	$	13,266

N.B. Costs are in Australian dollars, from the Disease Expenditure in Australia 2015-16 report prepared by the Australian Institute of Health and Welfare [21]. Includes overall health care costs per person minus costs of diseases and injuries included in the model

Appendix G: Road transport casualties in NSW, 2018

Road transport casualties in NSW, 2018

	Striking Mode*									
Victim#	Car/car derivative	Light truck	Heavy Truck	Bus	Motorcycle	Pedal Cycle	Pedestrian	Other	Single mode Crash	Total
Car/car derivative	6357	1434	490	79	79	14	83	162	2925	11623
Light truck	643	248	97	15	16	3	24	19	643	1708
Heavy rigid truck	40	19	11	0	4	2	4	3	57	140
Articulated truck	16	9	24	0	1	1	1	0	119	171
Bus	41	11	6	7	0	0	7	1	27	100
Other motor vehicle	34	13	4	1	0	0	0	3	26	81
Motorcycle	898	151	32	14	35	3	13	32	953	2131
Pedal cycle	462	84	17	5	5	19	9	44	72	717
Non-motorised vehicle	0	0	0	0	0	0	0	0	0	0
Pedestrian	852	149	23	25	25	11	0	86	0	1171
Other or unknown	1	0	0	0	0	0	0	1	1	3
Total	9344	2118	704	146	165	53	141	351	4823	17845

Victim - Occupant Casualty from Traffic Unit Type where Vehicle is the Key Vehicle or Other Vehicle in the First Impact

* Striking Mode - Traffic Type involved in the impact with the victim traffic unit type

Road transport fatalities in NSW, 2018

	Striking Mode*									
Victim#	Car/car derivative	Light truck	Heavy Truck	Bus	Motorcycle	Pedal Cycle	Pedestrian	Other	Single mode Crash	Total
Car/car derivative	43	16	21	2	0	0	0	1	87	170
Light truck	0	4	4	1	0	0	0	1	19	29
Heavy rigid truck	0	1	0	0	0	0	0	0	1	2
Articulated truck	0	0	2	0	0	0	0	0	5	7
Bus	0	0	0	0	0	0	0	0	0	0
Other motor vehicle	0	1	0	0	0	0	0	0	3	4
Motorcycle	12	5	3	1	2	0	0	0	31	54
Pedal cycle	2	1	2	0	0	0	1	0	3	9
Non-motorised vehicle	0	0	0	0	0	0	0	0	0	0
Pedestrian	39	15	5	3	2	0	0	3	0	67
Other or unknown	0	0	0	0	0	0	0	0	0	0
Total	96	43	37	7	4	0	1	5	149	342

Victim - Occupant Casualty from Traffic Unit Type where Vehicle is the Key Vehicle or Other Vehicle in the First Impact

* Striking Mode - Traffic Type involved in the impact with the victim traffic unit type

Number of trips per capita at baseline (Source NSW household survey)

Transport mode	Total trips (per weekday)	Mode share
Private vehicle (driver and passenger)	19,978,208	73.6%
Walking	4,042,999	14.9%
Cycle	540,109	2.0%
Public Transport	2,570,703	9.5%
Total	27,132,019	

Appendix H: Criteria for appraisal of evidence

Criteria for appraisal of evidence, assessment for causal relationship and grading of evidence

Appraising the quality of the evidence presented in the systematic reviews

We used a modified version of GRADE (Grading of Recommendations, Assessment, Development, and Evaluation) guidelines [45].

We applied the 7 *criteria* and '*what to look for*' as described by Webb and colleagues [45] with expanded assessment on our areas of interest such as; measurement of exposure, measurement and definition of outcome, internal validity; random error, chance, by looking at CI ranges, p values, systematic errors, bias by reporting on selection bias, information bias, and confounding. Our appraisals were guided by the reported information in our included studies.

- Criteria
  i. Focused Research question
  ii. Inclusion and exclusion criteria
  iii. Comprehensiveness of search strategy
  iv. Assessment of included studies
  v. Reproducibility of assessments
  vi. Similarity of results of include studies
  vii. Overall logic and insight

Assessing the reviews against causal criteria

We followed the Bradford Hill criteria [45, 46]. A summarised version of the criteria adopted is outlined below

Criterion	Description
1.Temporality	A cause must precede the effect in time. Observation in which cause (exposure) followed effect (outcome) merely shows that the exposure could not have caused the outcome in that instance; it provides no evidence for or against the hypothesis that the exposure can cause the outcome in instances which the exposure precedes the outcome. Only if found that the exposure cannot precede the outcome, can one dispense the causal hypothesis that exposure could cause the outcome.
2. Strength of association	We assessed measures as described by the relative effect, odds ratios or relative risks. We used an adaptation of measures by Webb and colleagues [45] to guide our classification. RR > 3.0 (<0.33), moderately strong; > 5.0 (<0.2), strong. *Additional measures:* RR 1.5-2.9 (0.34-0.67), modest; RR <1.5 (>0.67), weak association.

	A strong association is neither necessary nor sufficient for causality, weakness is neither necessary nor sufficient for absence of causality.
3. Consistency	Repeated observation of an association from other studies in different populations under different circumstances. Lack of consistency does not rule out a causal association. The effect of a causal agent cannot occur unless the complementary component causes act to complete a sufficient cause. In some circumstances, these conditions may not be met.
4. Dose-response relationships	A dose-response relationship can add weight to an evaluation of causation, but its absence need not count against a causal link. Relationship not always linear.
5.Biological plausibility	If there is a likely biological mechanism through which an exposure might cause the disease. This can add substantial weight to a casual argument. Lack of plausibility does not necessarily rule out causation, because increasing knowledge of disease mechanisms may reveal an association to be more credible with time.
6. Specificity	Where the relationship between exposure and disease is specific. A cause leads to a single effect, not multiple effects, and that an effect has one cause, not multiple causes. This comes into play when it can logically be deduced from the causal hypothesis in question and when non-specificity can be logically deduced from one or more noncausal hypotheses.
7.Coherence	A cause-and-effect interpretation for an association does not conflict with what is known of the natural history and biology of the disease. The absence of coherent information should not be taken as evidence against an association being considered causal. The presence of conflicting information may refute a hypothesis, but the conflicting information may be mistaken or misinterpreted.

Grading the evidence

We graded the evidence to support a judgement of a relationship. This grading process was guided by the World Cancer Research Fund grading system [23] with modification to best align to the nature of evidence that we assessed in our systematic review of reviews. A modification of the WCRF grading system was applied in the GBD 2017 study [22]. We adopted grades of convincing, probable, possible and insufficient evidence.

Convincing (strong) evidence

Evidence strong enough to support a judgement of a convincing causal (or protective) relationship, which justifies making recommendations designed to reduce the risk of health outcome. All the following are generally required:

- Evidence from at least two independent cohort studies

- No substantial unexplained heterogeneity within or between study types or in different populations relating to the presence or absence of an association, or direction of effect.
- Good quality studies to exclude with confidence the possibility that the observed association results from random or systematic error, inducing confounding, measurement error and selection bias.
- Evidence showing consistent associations between exposure and disease, with little or no evidence to the contrary
- The association should be biologically plausible.
- Presence of a plausible biological gradient ('dose-response') in the association. Such a gradient need not be linear or even in the same direction across the different levels of exposure, so long as this can be explained plausibly.

Probable evidence

Evidence strong enough to support a judgement of a probable causal (or protective) relationship, which generally justifies recommendations designed to reduce the risk of health outcome.

Evidence based on epidemiological studies showing fairly consistent associations between exposure and disease, but for which there are perceived shortcomings in the available evidence or some evidence to the contrary, which precludes a more definite judgment. Shortcomings in the evidence may be any of the following: insufficient duration of or studies; insufficient or studies available; inadequate sample sizes; or incomplete follow-up.

All the following are generally required:

- Evidence from at least two independent cohort studies
- No substantial unexplained heterogeneity within or between study types or in different populations relating to the presence or absence of an association, or direction of effect.
- Good quality studies to exclude with confidence the possibility that the observed association results from random or systematic error, inducing confounding, measurement error and selection bias.
- The association should be biologically plausible.
- Presence of a plausible biological gradient ('dose-response') in the association. Such a gradient need not be linear or even in the same direction across the different levels of exposure, so long as this can be explained plausibly.

Possible (suggestive) evidence

Evidence that is too limited to permit a probable or convincing causal judgement but is suggestive of a direction of effect. The evidence may be limited in amount or by methodological flaws but shows a generally consistent direction of effect. This judgement is broad and includes associations whether the evidence falls only slightly below that required to infer a probably causal association through to those where the evidence is only marginally strong enough to identify a direction of effect. This judgement is very rarely sufficient to justify recommendations designed to reduce the risk of health outcome; any exception to this require special, explicit justification.

All the following are generally required:

- Evidence from at least two independent cohort studies
- The direction of effect is generally consistent though some unexplained heterogeneity may be present
- Evidence for biological plausibility

Insufficient evidence

Evidence based on findings of a few studies which are suggestive, but insufficient to establish an association between exposure and disease. More well-designed research is needed to support the tentative association.

Evidence is so limited that no firm conclusion can be made. This judgement represents an entry level and is intended to allow any exposure for which there are sufficient data to warrant consideration, but where insufficient evidence exists to permit a more definitive grading. The evidence may be limited by the amount of evidence in terms of the number of studies available, by inconsistency of effect, by methodological flaws (for example, lack of adjustment for known confounders) or by any combination of these factors.

This grading does not necessarily mean a judgement that there is evidence of no relationship. With further good quality research, any exposure graded in this way might in the future be shown to increase or decrease the risk of the health outcome under investigation.

Appendix I: Measures of association: physical activity and low back pain and osteoarthritis

Study	Findings
Low back pain	
Alzahrani, et al., [51]	• Compared to low level total physical activity, medium level total physical activity[a] was significantly associated with a decreased risk of developing low back pain. Fully adjusted risk ratios, 0.90, 95% CI 0.85 to 0.96 (n=7) [*.] • Compared to low level total physical activity, high level total physical activity[b] was not associated with low back pain. Pooled 1.00, 95% CI 0.92 to 1.08 (n=9) [*.]
Shiri & Falah-Hassani, 2017 [52]	• Leisure time physical activity (LTPA) was neither associated with low back pain in the past month nor associated with low back pain in the past 6–12 months LTPA and Low back pain in the past month Active[c] Vs inactive (n=5) RR 0.98 95% CI 0.84 to 1.14 Moderate[d] Vs low activity (n=3) RR 0.93 95% CI 0.80 to 1.08 High[e] Vs Low activity (n=4) RR 0.85 95% CI 0.53 to 1.37 LTPA and Low back pain in the past 6–12 months Active[c] Vs inactive (n=16) RR 0.97, 95% CI 0.92 to 1.01 Moderate[d] Vs low activity (n=10) RR 0.99, 95% CI 0.87 to 1.07 High[e] Vs Low activity (n=11) RR 0.90, 95% CI 0.77 to 1.05 High[e] Vs Low or moderate[d] activity (n= 5) RR 1.01, 95% CI 0.89 to 1.15 • LTPA was associated with frequent or chronic low back pain as per adjusted risk ratios below. Active[c] Vs inactive (n=9) RR 0.84, 95% CI 0.77 to 0.92 Moderate[d] Vs low activity (n=3) RR 0.86, 95% CI 0.79 to 0.94 High[e] Vs Low activity (n=3) RR 0.70, 95% CI 0.48 to 1.03 • Adjusted RRs for associations between physical activity and frequent or chronic low back pain (Ref group: individuals with no regular physical activity) physically active individuals RR=0.89, 95% CI 0.82 to 0.97 (n=6) moderately active individuals RR=0.86, 95% CI 0.79 to 0.94 (n=2) highly active individuals RR=0.84, 95% CI 0.75 to 0.93 (n=2)

[a] Between the 33th and 66th percentile (reporting >11.89 and <21 MET-hours/week).

[b] Greater than or equal to the 66th percentile (reporting ≥21 MET-hours/week)

[*] Results from included cohort studies

[c] being physically active was defined as participation in a sport or other physical activity during leisure time, at least 1–2 times a week, at least 0.5–1.0 hour per week, or being in the middle or upper third of the distribution of leisure time physical activity in a study sample.

[d] Moderate level of physical activity was defined as participation in such activity 1–3 times a week,..1–3 hours per week..or being in the middle third of the distribution of leisure time physical activity in a study sample.

[e] High level of physical activity was defined as participation in leisure time physical activity, ≥ 3–4 times per week,..more than 2–4 hours per week...or being in the upper third of the distribution of such activity in a study sample

Abbreviations: RR, risk ratios; CI, confidence interval

Appendix J: Measures of association: Physical activity and all-cause mortality

Author, Date: health outcome	**Findings**
All-cause mortality	
Ekelund et al., 2019 [30]	• Total physical activity[a] was associated with lower risk of mortality with a non-linear dose-response. Hazard ratios, 1.00 (referent) in the first quarter (least active); second quarter[b], 0.48 (95% CI, 0.43 to 0.54); third quarter[c], 0.34 (95% CI, 0.26 to 0.45); fourth quarter[d] (most active), 0.27 (95% CI, 0.23 to 0.32) (n=8).
Hamer & Chida, 2008 [58]	• The pooled hazard ratio for ACM and the highest walking category[e] compared with the lowest was 0.68 (95% CI, 0.59 to 0.78) (n=10). • The pooled hazard ratio for ACM and minimal walking levels[f] compared with the referent category (using the same studies as for the main analyses), was 0.80 (0.71 to 0.91) (n=10).
Hupin et al., 2015 [59]	• Multivariate-adjusted all-cause mortality relative risks (RR) for participants of the low, medium and high moderate-to-vigorous intensity physical activity (MVPA) groups compared with participants in the inactive group were reported as below (n=9); Low-dose (1–499 MET-min per week) • RR 0.78, 95% CI 0.71 to 0.87 Medium dose [150 min of MVPA (500–999 MET-min) per week] • RR 0.72, 95% CI 0.65 to 0.80 High-dose (well above current recommendations, ≥1000 MET-min/wk) • RR 0.65, 95% CI 0.61 to 0.70 •
Kelly et al., 2014 [60]	• An association between 11.25 MET hours/week of walking and ACM was reported, Relative risk, 0.89 (95% CI 0.83 to 0.96) (n=14) • An association between 11.25 MET hours/week of cycling and ACM was reported, relative risk, 0.90 (95% CI 0.87 to 0.94) (n=7)
Lollgen, Bockenhoff, & Knapp, 2009 [61]	• Reported results of the multivariate-adjusted relative risk estimates for both sexes, for studies with three levels of activity were; moderate levels, 0.78 (95% CI, 0.61 – 1.00), and vigorous levels, 0.80 (95% CI, 0.66 – 0.97) (n=3). Ref group was inactive.
Nocon et al., 2008 [62]	• Relative risk of all-cause mortality in physically active versus physically inactive participants for fully adjusted models, were reported as 0.67 (95% CI 0.63, 0.72) (n=33). •

<table>
<tr><td>Samitz, Egger, & Zwahlen, 2011 [63]</td><td>

Results on total physical activity (leisure time PA, routine activities of daily living and occupational PA) were; mortality from all causes in individuals with highest compared with lowest levels of total physical activity, adjusted RRs, 0.65 (95% CI 0.60-0.71) (n=21)
Below are overall, maximally adjusted RRs of all-cause mortality associated with an increment of 2, 4 and 7 MET-h in total physical activity per day compared with lowest level of activity (the median level for the lowest activity category was 27.3 MET-h/day; twenty-four MET-hours correspond to sitting quietly for 24 h) (n=6).

<table>
<tr><td>2 MET-h/day (~850 MET-min/week)</td><td>0.95 (95% CI, 0.93–0.96)</td></tr>
<tr><td>4 MET-h/day (~1800 MET-min/week)</td><td>0.90 (95% CI, 0.87–0.92)</td></tr>
<tr><td>7 MET-h/day (~3000 MET-min/week)</td><td>0.83 (95% CI, 0.79–0.87)</td></tr>
</table>

</td></tr>
<tr><td>Woodcock, Franco, Orsini, & Roberts, 2011 [64]</td><td>

The association between hours of moderate physical activity and relative risk for all-cause mortality was reported as follows;

<table>
<tr><th>Hours/week</th><th>Moderate activity*</th></tr>
<tr><td>0</td><td>1.00 (Referent)</td></tr>
<tr><td>1</td><td>0.84 (95% CI, 0.81-0.88)</td></tr>
<tr><td>2.5</td><td>0.81 (95% CI, 0.76-0.85)</td></tr>
<tr><td>5</td><td>0.77 (95% CI, 0.73-0.82)</td></tr>
<tr><td>7</td><td>0.76 (95% CI, 0.71-0.81)</td></tr>
<tr><td>10</td><td>0.74 (95% CI, 0.68-0.79)</td></tr>
<tr><td>14</td><td>0.72 (95% CI, 0.66-0.78)</td></tr>
</table>
</td></tr>
</table>

[a] Total physical activity measured in counts per minute (cpm). [b] prespecified knot at the 25th centile of the exposure variable distributions using the medians of the quarters to define the exposure levels = 168cpm. [c] prespecified knot at the 50th centile of the exposure variable distributions using the medians of the quarters to define the exposure levels =256cpm. [d] prespecified knot at the 75th centile of the exposure variable distributions using the medians of the quarters to define the exposure levels =335cpm.

[e]highest walking exposure groups averaged more than 5.2 hours per week or more than 17.2 km per week- with considerable variation between studies.

[f]average walking time/distance in the minimal walking categories = approximately 3 hours per week (ranging from ,30 minutes per week to ,5 hours per week) or 9.8 km per week (ranging from ,5 km per week to ,15 km per week), which equated to a casual or moderate walking pace of approximately 3 km per hour.

*n activity of approximately 4.5 METs such as walking at 5.6 kilometres per hour carrying less than 11 kilograms

Abbreviations: CI, confidence interval

Appendix K: Physical activity and mental health

Introduction

The aim of this sub-study was to assess whether physical activity improves mental health, specifically depressive and anxiety disorders. We performed a systematic review of systematic reviews and applied criteria for causal inference to assess the likelihood of an association being causal.

Methods

Data Sources and search strategy

The search strategy was informed by guidelines for systematic reviews and rapid systematic reviews [95]. The review was prepared according to the Preferred Reporting Items for Systematic reviews and Meta-Analysis Protocols (PRISMA-P) 2015 statement [96]. The reviews covered the two concept areas: exposure and health outcome. The search was restricted to studies carried out on humans.

The development of the search strategies followed an iterative process. The search was informed by key articles identified in Phase 1 of the project and consultation of a research librarian at Griffith University. The search strategy was tested and modified in PubMed and Embase. We also reviewed reference lists from included studies for suitable studies that met the inclusion criteria. We studied the medical subject heading search (MeSH) terminology in PubMed to inform the selection of our search terms. Appendix table K-1 shows the search strategy that was used.

Appendix table K-1 Search strategies for the systematic review

Exposure and health outcome search terms
(“physical activity” OR “physical exercise” OR walking OR bicycling OR “active transport”) AND (“mental health” OR “mental disorders” OR anxiety OR “anxiety disorder*” OR depression OR “depressive disorder*”)

Inclusion and exclusion criteria

The inclusion and exclusion criteria described below were agreed upon by all reviewers.

- Participants/ Population: the review included studies reporting results for whole populations who did not have any of the study outcomes at baseline, covered either all ages, multiple age categories, or specific age groups so long as they are representative of the whole population at that age
- Studies reporting findings of the associations between PA (exposure) and the outcomes: depression, anxiety.
- Study designs: systematic reviews and meta-analyses
- Publication status: published studies or studies in print whose full text is publicly available. Most recent versions were considered for studies published in multiple papers. Duplicate publications of the same material were excluded.
- Timeline: Studies published in the year 2000 to 19/3/2020
- Language: English
- Studies that provided risk estimates (relative risk, hazard ratio, or odds ratio) with confidence intervals or standard errors, or the data needed to calculate them were included.

Study Records

Data management

The identified studies were imported to EndNote X9 software. Duplicate records were identified and excluded. The inclusion and exclusion criteria above guided our study selection. The full text of the selected articles was retrieved and saved into EndNote X9 software.

Screening

One reviewer (MW) screened the titles and abstracts of identified studies for relevance based on the set criteria. One reviewer then screened the full texts of studies that were potentially eligible for inclusion. The other authors reviewed and confirmed the selection process. The final list of studies was discussed and agreed upon by all authors. The reasons for exclusion of studies after full text review are documented in Appendix N.

Data extraction

All reviewers achieved consensus on which data to extract from included studies. One reviewer (MW) extracted data from the full texts of eligible studies. Two other reviewers (HM and LV) cross-checked the data extraction variables. The main areas of data extraction from each study meeting the inclusion criteria are shown in Box 1.

Box 1 Data extraction fields systematic review

- 1st author's last name
- Year of data analysis - dates in search criteria
- Year of publication
- AMSTAR score
- Aim
- Study design selection
- Population
- Reported quality of evidence in the included reviews
- Exposure Details
- Outcome Details
- Data analysis applied
- Effect size reported
- Dose-response results reported
- Publication bias assessed
- Published results of sub-group analyses
- Main conclusion
- Funding
- Other comments

Risk of bias and quality appraisal

We assessed the methodological quality of the included reviews using the Assessment of Multiple Systematic Reviews (AMSTAR) rating scale [97].

Data Synthesis

We identified the most up to date studies that quantified the relationship between physical activity and risks of depression, anxiety. Study selection was guided by the quality scores for each study, presentation of elaborate details on exposure, outcome, and association measures presented in the papers. Alternative reported values were identified for sensitivity analyses in our model. Our criteria for appraisal of evidence, assessment for causal relationship and grading of evidence is presented in Appendix table K-2.

Appendix table K-2 Criteria for appraisal of evidence, assessment for causal relationship and grading of evidence

Appraising the quality of the evidence presented in the systematic reviews

We used a modified version of GRADE (Grading of Recommendations, Assessment, Development, and Evaluation) guidelines [45].

We applied the 7 *criteria* and '*what to look for'* as described by Webb and colleagues [45] with expanded assessment on our areas of interest such as; measurement of exposure, measurement and definition of outcome, internal validity; random error, chance, by looking at CI ranges, p values, systematic errors, bias by reporting on selection bias, information bias, and confounding. Our appraisals were guided by the reported information in our included studies.

- Criteria
  - viii. Focused Research question
  - ix. Inclusion and exclusion criteria
  - x. Comprehensiveness of search strategy
  - xi. Assessment of included studies
  - xii. Reproducibility of assessments
  - xiii. Similarity of results of include studies
  - xiv. Overall logic and insight

Assessing the reviews against causal criteria

We followed the Bradford Hill criteria [45, 46]. A summarised version of the criteria adopted is outlined below

Criterion	Description
1.Temporality	A cause must precede the effect in time. Observation in which cause (exposure) followed effect (outcome) merely shows that the exposure could not have caused the outcome in that instance; it provides no evidence for or against the hypothesis that the exposure can cause the outcome in instances which the exposure precedes the outcome. Only if found that the exposure cannot precede the outcome, can one dispense the causal hypothesis that exposure could cause the outcome.

2. Strength of association	We assessed measures as described by the relative effect, odds ratios or relative risks. We used an adaptation of measures by Webb and colleagues [45] to guide our classification. RR > 3.0 (<0.33), moderately strong; > 5.0 (<0.2), strong. *Additional measures:* RR 1.5-2.9 (0.34-0.67), modest; RR <1.5 (>0.67), weak association. A strong association is neither necessary nor sufficient for causality, weakness is neither necessary nor sufficient for absence of causality.
3. Consistency	Repeated observation of an association from other studies in different populations under different circumstances. Lack of consistency does not rule out a causal association. The effect of a causal agent cannot occur unless the complementary component causes act to complete a sufficient cause. In some circumstances, these conditions may not be met.
4. Dose-response relationships	A dose-response relationship can add weight to an evaluation of causation, but its absence need not count against a causal link. Relationship not always linear.
5.Biological plausibility	If there is a likely biological mechanism through which an exposure might cause the disease. This can add substantial weight to a casual argument. Lack of plausibility does not necessarily rule out causation, because increasing knowledge of disease mechanisms may reveal an association to be more credible with time.
6. Specificity	Where the relationship between exposure and disease is specific. A cause leads to a single effect, not multiple effects, and that an effect has one cause, not multiple causes. This comes into play when it can logically be deduced from the causal hypothesis in question and when non-specificity can be logically deduced from one or more noncausal hypotheses.
7.Coherence	A cause-and-effect interpretation for an association does not conflict with what is known of the natural history and biology of the disease. The absence of coherent information should not be taken as evidence against an association being considered causal. The presence of conflicting information may refute a hypothesis, but the conflicting information may be mistaken or misinterpreted.

Grading the evidence

We graded the evidence to support a judgement of a relationship. This grading process was guided by the World Cancer Research Fund grading system [23] with modification to best align to the nature of evidence that we assessed in our systematic review of reviews. A modification of the WCRF grading system was applied in the GBD 2017 study [22]. We adopted grades of convincing, probable, possible and insufficient evidence.

Convincing (strong) evidence

Evidence strong enough to support a judgement of a convincing causal (or protective) relationship, which justifies making recommendations designed to reduce the risk of health outcome. All the following are generally required:

- Evidence from at least two independent cohort studies
- No substantial unexplained heterogeneity within or between study types or in different populations relating to the presence or absence of an association, or direction of effect.
- Good quality studies to exclude with confidence the possibility that the observed association results from random or systematic error, inducing confounding, measurement error and selection bias.
- Evidence showing consistent associations between exposure and disease, with little or no evidence to the contrary
- The association should be biologically plausible.
- Presence of a plausible biological gradient ('dose-response') in the association. Such a gradient need not be linear or even in the same direction across the different levels of exposure, so long as this can be explained plausibly.

Probable evidence

Evidence strong enough to support a judgement of a probable causal (or protective) relationship, which generally justifies recommendations designed to reduce the risk of health outcome.

Evidence based on epidemiological studies showing fairly consistent associations between exposure and disease, but for which there are perceived shortcomings in the available evidence or some evidence to the contrary, which precludes a more definite judgment. Shortcomings in the evidence may be any of the following: insufficient duration of or studies; insufficient or studies available; inadequate sample sizes; or incomplete follow-up.

All the following are generally required:

- Evidence from at least two independent cohort studies
- No substantial unexplained heterogeneity within or between study types or in different populations relating to the presence or absence of an association, or direction of effect.
- Good quality studies to exclude with confidence the possibility that the observed association results from random or systematic error, inducing confounding, measurement error and selection bias.
- The association should be biologically plausible.
- Presence of a plausible biological gradient ('dose-response') in the association. Such a gradient need not be linear or even in the same direction across the different levels of exposure, so long as this can be explained plausibly.

Possible (suggestive) evidence

Evidence that is too limited to permit a probable or convincing causal judgement but is suggestive of a direction of effect. The evidence may be limited in amount or by methodological flaws but shows a generally consistent direction of effect. This judgement is broad and includes associations whether the evidence falls only slightly below that required to infer a probably causal association through to those where the evidence is only marginally strong enough to identify a direction of

effect. This judgement is very rarely sufficient to justify recommendations designed to reduce the risk of health outcome; any exception to this require special, explicit justification.

All the following are generally required:

- Evidence from at least two independent cohort studies
- The direction of effect is generally consistent though some unexplained heterogeneity may be present
- Evidence for biological plausibility

Insufficient evidence

Evidence based on findings of a few studies which are suggestive, but insufficient to establish an association between exposure and disease. More well-designed research is needed to support the tentative association.

Evidence is so limited that no firm conclusion can be made. This judgement represents an entry level and is intended to allow any exposure for which there are sufficient data to warrant consideration, but where insufficient evidence exists to permit a more definitive grading. The evidence may be limited by the amount of evidence in terms of the number of studies available, by inconsistency of effect, by methodological flaws (for example, lack of adjustment for known confounders) or by any combination of these factors.

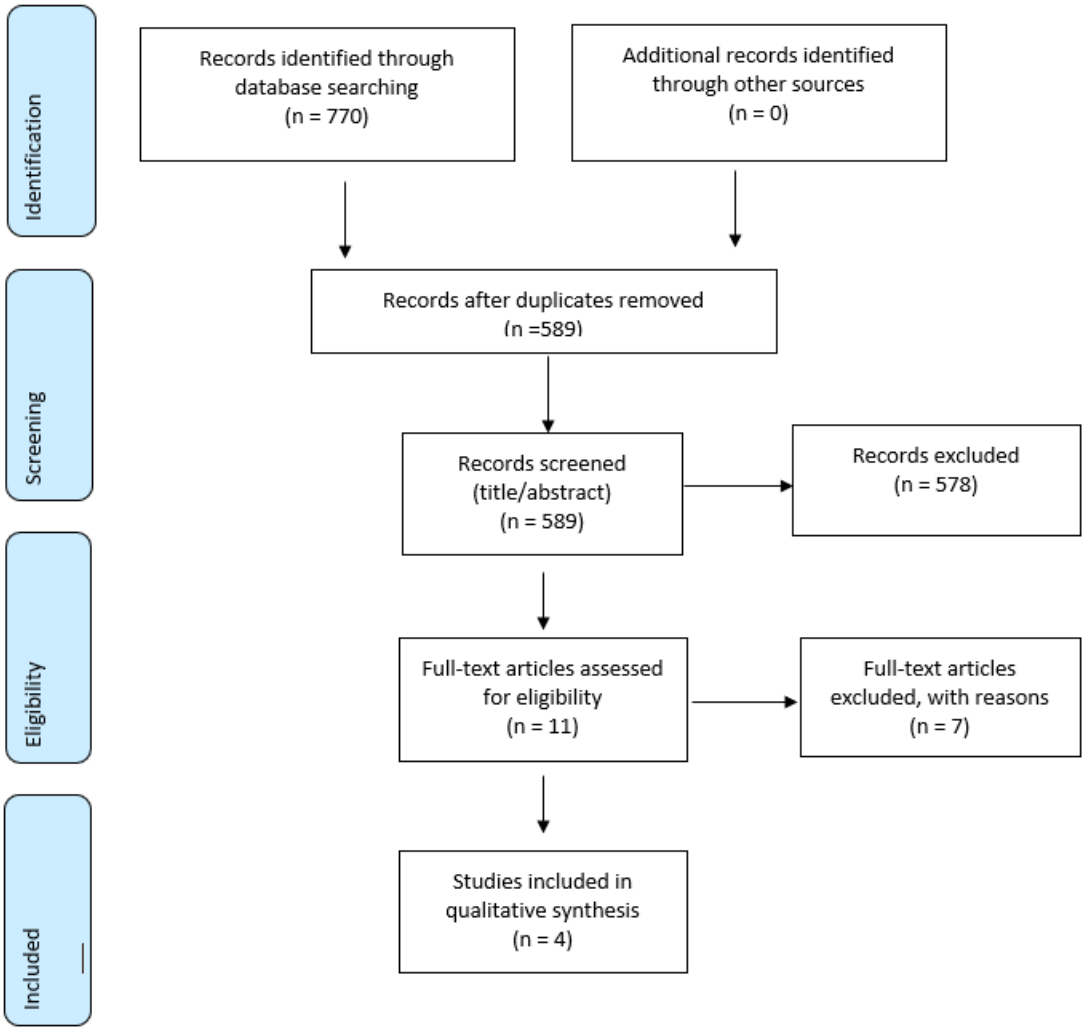
This grading does not necessarily mean a judgement that there is evidence of no relationship. With further good quality research, any exposure graded in this way might in the future be shown to increase or decrease the risk of the health outcome under investigation.

Results

Literature database search

The total number of articles identified is presented in the PRISMA diagram (Appendix Figure K-1). The database search identified 770 articles. A total of 589 records remained after removing duplicates. After screening of title and abstract, 578 records were excluded and 11 remained for full text analysis. A further 7 studies were excluded with reason after full text analysis leaving a total of 4 studies [47-50] for data extraction. Reasons for exclusion were; study investigated the association between physical activity and existing anxiety and depressive symptoms (n= 4), exposure variable in the study was not physical activity (n=1) and study did not describe an association or give measures of association between our selected exposure and outcome (n=1). An additional systematic review of reviews that included incident depression as an outcome of study was excluded. The authors included one review paper on incident depression. This paper had already been identified for inclusion in our qualitative synthesis. On this basis, this systematic review of reviews paper was excluded (n=1). Appendix N provides a list of the studies excluded after full-text analysis and specific reasons for exclusion. Appendix O summarizes the quality scores for each of the included studies. Using the AMSTAR criteria [97], two of the included studies were scored as high quality reviews [48, 50]. Of these high-quality studies, one represented the results for depression and another represented results

for anxiety. The other two studies (one representing results for depression and another for anxiety) were scored as moderate quality reviews [47, 49].



Appendix Figure K-1 Prisma flow diagram- Physical activity and health outcomes, depression and anxiety

Depression

Strength of association

The systematic reviews by Mammen and Faulkner [47] and Schuch and colleagues[48] provide evidence of an association between PA and depression. Mammen and Faulkner [47] perfomed a systematic review of prospective cohort studies. They found a significant, inverse relationship between baseline PA and depression in later years in 25 of the 30 included studies, after adjusting for potential other explanatory variables. This suggests that physical activity can prevent the onset of depression. A similar finding was seen in a meta-analysis of prospective cohort studies done by Schuch

and colleagues[48]. The authors found that compared with people with low levels of PA, those with higher levels of PA have lower risks of developing depression.

A summary of the measures of disease association found in our review study is presented in Appendix K. For the association between PA and depression, measures of association for use in our model were taken from the Schuch and colleagues[48] study (adjusted relative risk=0.83, 95% CI=0.76, 0.90). For modelling purposes, we assumed that the *lowest* category in their paper refers to the *inactive* category in our model, and *highest* category refers to the *highly active* category in our model. Scaling was done to interpolate values between 'highest' and 'lowest' in the studies, using the a study by Ekelund and colleagues [30] which pooled the results of accelerometry-measured PA (see section 4.2.5). Appendix table K-3 presents a summary of findings on the Association between physical activity and depression.

Appendix table K-3 Summary of findings on the Association between physical activity and depression

Author, Date: sub-divided by health outcome	Findings
Depression	
Mammen & Faulkner, 2013 [47]	• Engaging in <150 minutes/week was associated with 8%–63% decreased risk of future depression (n=3) • Engaging in >150 minutes/week was associated with 19%–27% decreased risk of future depression (n=3)
Schuch et al., 2018 [48]	• Compared with people with low levels of physical activity, those with high levels had decreased risks on adjusted relative risk analyses: adjusted relative risk, 0.83, 95% CI=0.76, 0.90 (n=18)

Appraisal of evidence and assessment of causal relationship

Two studies that investigated the association between exposure to PA and depression met the inclusion criteria of our systematic review [47, 48].

In a systematic review of 30 prospective studies, Mammen and Faulkner [47] investigated whether PA was protective against the onset of depression. The authors included prospective-based, longitudinal design articles examining relationships between PA and depression over at least two-time intervals. Only studies that excluded individuals with depression at baseline were considered. The studies' follow up period ranged from 1 to 27 years.

In the studies included in this review, subjective PA measures of aerobic activity were used with only one study reporting objectively measured PA via ergometer cycling. The use of (imprecise) self-report measures for PA and depression in the reviewed studies may also have introduced 'regression dilution bias' and an increase in random variation, respectively, which might conceal true associations between PA and depression. The authors reported that the majority of the studies assessed depression through well-validated measures.

Two reviewers independently conducted a formal quality assessment for each study using the Critical Appraisal Skills Programme (CASP) for prospective studies. The CASP protocol appraises both internal

and external validity by addressing four methodologic issues: (1) selection bias; (2) PA measurement bias; (3) depression measure bias; and (4) accounted confounding variables. Out of the 30 studies included in their analyses, 25 found that baseline PA was negatively associated with a risk of subsequent depression. Most studies that found a protective role were of high (n=17) or modest (n=6) methodologic quality. The authors propose a cautious interpretation of the study results since several covariates, such as genetic variations that predict both PA and depression, were not fully accounted for. Additionally, the authors suggested that other sources of literature not included in the review; such as grey research, conference papers, research reports, Ph.D. dissertations, and studies not reported in English, may have contributed to publication bias. Hence, the positive findings may be a consequence of publication bias. Nevertheless, the review was conducted comprehensively, and pertinent scholarly articles identified from various databases.

The use of inconsistent self-report measures of PA between studies prevented Mammen and Faulkner [47] from examining the dose-response relationships between PA and depression. However, these authors presented a narrative overview of findings from 7 studies that reported dose-response relationships. Of the seven studies, five were of high quality. For instance, one high-quality study found that as little as 10–29 minutes (RR=0.90) of daily PA was preventive in the onset of depression. Higher levels of daily PA were associated with further decreases in the risk of developing depression (60–90 minutes/day, RR=0.84; >90 minutes/day, RR=0.80) [47].

Four of the studies reviewed concluded that the protective effect of PA on depression was specific to women and girls. These studies suggest that psychological factors may explain these findings because women may benefit more from the social aspects of PA than men. Of the five studies that revealed null findings for the association between PA and depression, only one was considered high quality. The low and modest scores assigned to these studies were due to the failure to control for significant covariates that have been linked with depression, such as body mass and socio-economic status. The review authors discuss that in four of these studies, the lack of precision of estimating PA dosage might have explained the null findings.

For the protective effect of PA on depression, no single mechanism was considered to fully explain the preventive effect observed. The authors proposed that a range of physiologic, biochemical, and psychosocial mechanisms most likely operate together. The authors concluded that their findings presented sufficient evidence to conclude that any level of PA could prevent future depression.

In the second study included in the qualitative synthesis, Schuch and colleagues [48] carried out a meta-analysis of prospective cohort studies to examine the prospective relationship between PA and incident depression. Unlike the study by Mammen and Faulkner [47], which considered individual study results separately, the pooled meta-analysis in the Schuch study allows a clearer understanding of a true association between PA and depression.

Only prospective design studies with at least 1 year of follow-up were included. Less than one year was considered an insufficient time frame for risk and protective factors to exert a meaningful influence on depressive symptoms. A total of 49 studies were included with an average follow up time of 7.4 years across the studies. The majority of the included studies provided data for adjusted measures of association meaning that various covariates were considered in the primary studies. The mean study quality score of the studies was 6.34 (SD=0.8) out of 9, representing moderate to high methodological quality.

The authors found that compared with people with low levels of PA, those with higher levels of PA are consistently associated with reduced risks of developing depression (adjusted relative risk=0.83, 95% CI=0.76, 0.90). The authors found that the protective effect seen was for people of all ages (youths,

working-age adults, elderly persons). These results were consistent with and without adjustment for possible confounding factors.

In this meta-analysis, only one of the included studies used an objective measure (pedometer) to evaluate PA. The rest used self-report questionnaires to measure PA. Fifteen studies evaluated major depression using structured or semi-structured diagnostic instruments or self-reported physician diagnosis of depression. This limitation of self-reported data was also seen in the study by Mammen and Faulkner [47]. Self-report questionnaires are associated with bias such as recall bias. However, in this study subgroup analyses showed that PA decreased the risk of developing depression, regardless of whether this was based on self-report measures or major depression diagnosis from structured clinical diagnostic interviews.

The authors mention that despite including only studies in which there were no depressed participants at baseline, the risk of selection bias was not entirely excluded. This is largely because depression is a recurrent disorder and previous depressive episodes were not well documented in the studies that they investigated. The authors also reported that seven of their subgroup analyses were nonsignificant. This was attributed to the small number of studies included in some of the subgroup analyses. Despite their robust findings, the authors warn that caution is required as some covariates may not have been assessed in this study. Examples of such covariates are genetic factors, familial history of depression, other risk factors for depression, such as obesity, poor diet, and use of tobacco, and other clinical comorbidities.

The differences in the assessment of depressive symptoms at baseline across studies were also seen as a limitation. The authors reported that it may have been possible that the inclusion of participants who exhibited subthreshold depressive symptoms at baseline influenced the likelihood of developing depression at follow-up because of either a lower engagement in PA or because of an inherently higher risk of developing full-blown depression. Nonetheless, included studies that controlled for baseline depressive symptom severity in a subgroup analysis also reported significant associations between high PA levels and lower development of depression. This showed that a protective effect of PA was also evident in people with subthreshold depressive symptoms.

Unlike Mammen and Faulkner [47] who reported that gender may modify the effect of PA on incident depression, the findings by Schuch and colleagues [48] suggested that the potential protective association of PA is similar for men and women. In their subgroup analyses, their findings demonstrated that the protective effects of physical activity were found in studies in which the different aspects of PA (intensity, frequency, volume) were measured individually or when two or more aspects (metabolic equivalents/composite) were considered. Correction for publication bias, slightly reduced the associations, but the association remained significant (adjusted odds ratio=0.85; 95% CI=0.81, 0.89; adjusted relative risk=0.86; 95% CI=0.78, 0.96).

A major limitation in this study was that the definitions of low or high PA, as well as the aspects of PA that were captured by each instrument (intensity, frequency, volume, or two or more), varied widely between the primary studies included in the review. These limitations prevented the study from establishing a “minimum” or an “optimal” dosage of PA necessary to decrease the odds of incident depression. Nonetheless, the authors considered the evidence available to support the conclusion that PA can offer protection against the emergence of depression.

Similar to the discussion presented by Mammen and Faulkner [47], Schuch and colleagues [48] highlight that it is likely that no single mechanism can explain the protective effect of PA in the development of depression. The authors discuss that a range of biochemical and psychosocial factors are likely responsible, including biological mechanisms through which exercise increases neurogenesis

and reduces inflammatory and oxidant markers and activates the endocannabinoid system; a neuromodulatory system involved in several mental disorders. Moreover, they report that people with depression have decreased hippocampal volumes and levels of markers of neurogenesis, and increased levels of inflammatory and oxidant markers. There is evidence that PA may regulate these abnormalities, increasing hippocampal volume and neurogenesis levels, as well as adjusting the imbalance between anti- and proinflammatory and oxidant markers. Also, PA may directly increase psychological factors such as self-esteem or perceptions of physical competence. Improved levels of fitness lead to both subjective and objective improvements in physical health status. In both studies [47, 48], the authors recommend that future research should investigate these underlying biological and psychological mechanisms.

The section below presents a summary of the assessment against Bradford Hill's criteria for causality [45, 46] (Appendix K 4) The two review studies included primary results from prospective cohort studies. Inclusion of prospective based studies allowed for the examination of the temporal sequence between baseline levels of PA and primary prevention of depression at follow up. This gives additional insights by uncovering the direction of relationships which is one of Bradford Hill's main criteria for judging whether a causal link exists between two variables [45, 46] (Appendix H). Assessing prospective evidence is useful in making a stronger case for a causal link between PA and depression. We considered that the evidence from these studies meets the temporality criteria where the exposure (PA) is seen to precede the effect (depression).

From our classification of measures of effect in our guidelines (Appendix H), the relative effects reported from the studies were considered weak in association. But weakness of an association does not preclude causality. Moreover, consistency was observed across the various primary studies that had been carried out in different populations. A dose relationship was reported in various primary studies included in the review by Mammen and Faulkner [47]. Schuch and colleagues [48] did not investigate a dose response relationship due to the varied definitions of low or high PA seen in their included primary studies. As outlined in the Bradford Hill criteria for assessing causality, a dose-response relationship can add weight to an evaluation of causation [45, 46].

As discussed earlier, there are various possible biological mechanism through which insufficient PA could cause depression. We concluded that the available evidence meets the biological plausibility criteria.

In summary, the consistent findings from the two review studies included in our qualitative synthesis provide sufficient evidence that an association between PA and depression exists. When assessed against Bradford Hill's criteria for causality [45, 46], we graded our findings as ***probable*** evidence for a causal relationship. This supports the inclusion of depression in the *NSW Active Transport Health Model*.

Appendix table K-4 Evidence for causal relationship

Assessing the evidence against causal criteria: Depression	
Criterion	Findings
1.Temporality	Pooled results from prospective studies show that lower baseline levels of PA are associated with depression at follow up [47, 48].
2. Strength of association	From our classification of measures of effect in our guidelines (see Appendix H), the relative effects reported from the studies were considered weak. Weakness of an association makes the risk of alternative explanations greater but does not preclude causality.

3. Consistency	Consistent findings were observed across the two review studies included in our review. In their discussions, the authors of the two review studies also indicated that consistency had been witnessed in the various primary studies that had been carried out in different populations.
4. Dose-response relationship	A dose relationship was reported in various primary studies included in the review by Mammen and Faulkner [47]. A dose-response relationship can add weight to an evaluation of causation [45, 46]
5.Biological plausibility	There are various possible biological mechanisms through which insufficient PA could cause depression. We concluded that the available evidence meets the biological plausibility criteria.
6. Specificity	This criterion is not met – physical inactivity does not invariably lead to depression, and depression is not the only health condition associated with inactivity. However, this criterion was thought of in relationship to infectious agents, and seldom applies.
7.Coherence	We considered that the interpretation of the association between PA and depression does not conflict with what is known of the natural history and biology of depression.
Assessment of grade of evidence Convincing / Probable / Possible / Insufficient	When assessed against Bradford Hill's criteria for causality [45, 46], we graded the evidence as strong enough to support a judgement of a probable causal relationship, with higher levels of physical activity probably leading to a lower risk of depression. This supports the inclusion of depression in the *NSW Active Transport Health Model*.

Anxiety

Strength of association

In their recent systematic review and meta-analysis, McDowell and colleagues [49] synthesised population-based evidence of a prospective association between PA and incident anxiety disorders. The authors found that with higher PA exposure, the odds were significantly lower for self-reported anxiety symptoms, diagnosed anxiety disorder, and generalized anxiety disorder. The second study included in our qualitative synthesis was by Schuch and colleagues [50]. These authors performed a meta-analysis of prospective cohort studies to examine the relationship between PA and incident anxiety. They found that higher self-reported PA levels were associated with lower rates of incident anxiety when compared with lower PA levels. Appendix table K-5 presents a summary of the measures of disease association found in our review study.

For the association between PA and anxiety, we used the adjusted odds ratio from the strongest available meta-analysis by Schuch and colleagues [50] (AOR = 0.74, 95% CI = 0.62, 0.88). As this paper presented results for the comparison between highest and lowest exposure categories, the assumption was made to consider these as equivalent to the *inactive* category in the model and the *highly active* category, respectively. For the intermediate categories, scaling was done to interpolate values between 'highest' and 'lowest' in the studies, using the study by Ekelund and colleagues [30] which pooled the results of accelerometry-measured PA on mortality (see section 4.2.5).

Appendix table K-5 Summary of findings on the Association between physical activity and anxiety

Author, Date: sub-divided by health outcome	Findings
Anxiety	
McDowell, Dishman, Gordon, & Herring, 2019 [49]	• From the adjusted models, after physical activity exposure, the odds were significantly lower for; - Self-reported anxiety symptoms (AORs, 0.8742 [95% CI, 0.7731,0.9886]) (n=9), - diagnosis of any anxiety disorder (mean OR, 0.6626 [95% CI=0.5337, 0.8227]) (n=3), and - diagnosis of generalized anxiety disorder (mean OR, 0.5438 [95% CI=0.3231, 0.9153]) (n=3). •
Schuch et al., 2019 [50]	• Higher self-reported physical activity levels were associated with decreased incident anxiety when compared with lower PA levels, adjusted odds ratio, 0.74, 95% CI = 0.62, 0.88 (n=11). •

Appraisal of evidence and assessment of causal relationship

Two studies investigating the association between PA and anxiety met the inclusion criteria for our qualitative synthesis [49, 50]. Both were published in 2019. In their systematic review and meta-analysis, McDowell and colleagues [49] sought to synthesise population-based evidence of a prospective association between PA and incident anxiety disorders. A total of 24 prospective cohort studies were included in the systematic review. Of these 24, thirteen were included in the meta-analyses. The authors excluded all studies that had a follow up period from baseline as less than 1 year. They considered that <1 year was not a sufficient time frame for PA to exert a meaningful influence on incident anxiety disorders. However, few primary studies fully accounted for participants lost to follow-up. Nonetheless, inclusion of prospective based studies allowed for the examination of the temporal sequence between baseline levels of PA and primary prevention of anxiety at follow up. This is a main criterion for judging whether a causal link is likely to exist between PA and anxiety [45, 46].

The included studies used self-report measures of PA. Eighteen studies assessed PA at baseline alone, while six studies assessed PA levels at several points in time. In self-report measures, PA is often over-estimated leading to misclassification, for example, inactive people who were classified as active. In this study it is likely that misclassification may have resulted to an underestimation of the strength of the relationship between PA and anxiety. Additionally, most of the included studies limited PA exposure to a single measure at baseline. Only six studies measured PA more than once in their follow up. No studies regularly tracked physical activity levels in a cohort across time permit an estimate of change in exposure across follow-up. This would have reduced the risk of misclassification bias.

Anxiety was measured through varied measures such as self-reported anxiety symptoms, screening tests, self-reported diagnosis or a hospital diagnosis of an anxiety disorder. Though the described outcome measures in this study were considered high quality, a moderate degree of heterogeneity

was found for outcomes of self-reported anxiety symptoms and a diagnosis of any anxiety disorder. However, the small number of studies included in analyses did not allow for a quantitative examination (i.e., meta-regression analysis) of potential sources of variability.

Two authors independently assessed the quality of the included studies using the Q-Coh tool. This tool assesses seven domains which are derived from the classification of biases (selection bias, performance bias, detection bias and attrition bias). Specifically, the authors assessed the representativeness of the sample, comparability of the groups at the beginning of the study, quality of the exposure measure, maintenance of the comparability during the follow-up time, quality of the outcome measure, attrition, and statistical analyses. Six studies were assessed as low quality, nine as acceptable, and nine as good. From the authors' review, the key sources of bias in the primary studies included inconsistent adjustment for putative confounders, representativeness of samples and diversity in key sample characteristics, and attrition bias.

Results for the adjusted models showed that exposure to higher levels of PA was associated with a significantly lower risk of the occurrence of any anxiety disorder (OR=0.66, 95% CI, 0.53, 0.82, n=3) and generalised anxiety disorder (OR=0.54, 95% CI, 0.32, 0.92, n=3). Despite the challenges that the authors identified in the review study, they concluded that the evidence did suggest that engaging in PA protects against anxiety symptoms and disorders.

The second study included in our qualitative synthesis was a meta-analysis of prospective cohort studies conducted by Schuch and colleagues [50]. The authors examined the prospective relationship between PA and incident anxiety and sought to estimate the magnitude of the effect through a meta-analysis. A total of 14 cohorts of 13 unique prospective studies were included in their review. Only studies with a follow-up period of 1 year or longer were included. Similar to the McDowell study [49], Schuch and colleagues [48] also considered that less than 1-year follow-up may not have been a sufficient period for risk and protective factors to exert a meaningful influence on mental health symptoms.

Methodological quality was assesses using the Newcastle-Ottawa Scale (NOS). The NOS scale evaluates the risk of bias of prospective studies with three domains: (a) selection of participants (b) comparability of cohorts on the basis of the design of the analysis; and (c) outcomes. Selection of participants domain has four items assessing (a.1) representativeness of the exposed cohort, (a.2) same derivation between source of exposed and nonexposed participants, (a.3) ascertainment of the exposure, and (a.4) demonstration that the outcome of interest was not present at the baseline. The outcome domain has three items: (c.1) adequate assessment of the outcome of interest, (c.2) adequate duration of follow-up, and (c.3) adequacy of follow-up [50]. Overall study quality was moderate to high (mean NOS = 6.7 out of 9). This was suggestive for a low risk of bias of the included studies.

No study used an objective measure to evaluate PA hence may have been subject to measurement (recall) bias. The authors included studies that had evaluations of high versus low PA using any criterion. The PA classification criteria differed across the included studies. Also, only five of the included studies used validated PA measures and PA characterization differed across instruments used. The varied PA measures and classifications may have prevented the authors from assessing for a dose response relationship between PA and anxiety. Across the 14 cohorts, 10 studies evaluated anxiety disorders using structured or semi-structured diagnostic instruments or self-reported physician diagnosis of anxiety disorders, and 4 studies used cut-offs of anxiety screening instruments.

The authors noted that studies included in the meta-analysis did not specifically exclude people with incident depression from analysis. Depression is highly comorbid with anxiety and is also associated

with low PA. Still, the authors highlight previous evidence that shows that effects of exercise on anxiety are equal to and independent of effects of exercise on depression.

The results of this review indicate that people with high self-reported PA were at reduced odds of developing anxiety when compared with lower PA levels (AOR = 0.74, 95% CI = 0.62, 0.88). Also, completing the recommendation of 150 min of moderate/vigorous PA was associated with a lower risk of incident anxiety (AOR = 0.71, 95% CI = 0.54, 0.94).

Evidence of publication bias was encountered both for adjusted and crude analysis. Nonetheless, after adjusting for publication bias, the analyses in this study remained significant and did not change the overall effect size significantly (AOR = 0.86, 95% CI = 0.69, 0.99). The authors determined that the number of studies with negative results required to nullify the effects of PA on incident anxiety was 40 and 13, in AOR and OR analyses, respectively.

The authors included only 11 studies that provided effects adjusting for relevant covariates and six studies that provided crude OR. The covariates considered included age, sex, BMI, or combining age, sex, with BMI or smoking. Due to the small number of studies, the authors report that the results should be read with caution. Nevertheless, they conclude that the evidence from their study supports the notion that self-reported PA can confer protection against the emergence of anxiety.

This section presents a summary of the assessment against Bradford Hill's criteria for causality [45, 46] . For both studies by Schuch and colleagues [50] and McDowell and colleagues [49] studies, only prospective cohort studies, with a follow up period of more than 1 year, were included in the reviews. These studies assess prospective evidence where exposure (PA) preceded the effect (anxiety) in time a stronger case for a causal link between PA and anxiety can be made. We determined that the evidence meets the temporality criterion: exposure preceded outcome in the included cohort studies.

Assessing for causality based on strength of association, we compared the combined measures of association presented in both studies against our set criteria (Appendix 4). We used proposed measures by Webb and colleagues [45] to guide our classification of the strength of association between PA and anxiety. The mean OR was considered modest in strength though with a very wide CI. OR values >0.67 were classified as weak associations. Though the strength of association facilitates assessment for possible causal relationship, a strong association is neither necessary nor sufficient for causality, and weakness is neither necessary nor sufficient for concluding absence of causality.

In the study by McDowell and colleagues [49], all crude and adjusted associations included in their meta-analyses indicated inverse associations between physical activity and subsequent anxiety. Schuch and colleagues [50] provided additional evidence of the protective effects of self-reported PA on anxiety development referencing previous cross-sectional studies. We found that there was evidence for repeated observation of an association between PA and anxiety from other studies in different populations under different circumstances. Hence, we concluded that our findings present evidence that supports the criterion of consistency.

Another criterion for assessment of a causal relationship is presence of a dose-response relationship. A dose response relationship can add weight to an evaluation of causation, but its absence need not count against a causal link. As mentioned earlier in this section, the Schuch [50] study did not investigate a dose response relationship between physical activity exposure and anxiety. In the study by McDowell and colleagues [49] study, a total of 11 included studies assessed for a dose response relationship between PA and various anxiety outcomes. All the 11 studies reported lower odds of anxiety outcomes for increased amounts of PA. Of these 11 studies, six used a PA measure that

considered volume. However, different measures of exposure were used and criteria for classification into dose categories were not equivalent across studies. Only two studies reported associations for comparable dose categories (International Physical Activity Questionnaire categories).

We also assessed the evidence for biological plausibility to determine possible causal relationship. The presence of a likely biological mechanism through which an exposure might cause the disease can add substantial weight to a causal argument. Though the mechanisms are largely unclear, both studies presented evidence of potential biological processes that may underlie the protective effect of PA on incident anxiety. PA is known to influence similar pathways as those seen to play a role in the pathogenesis of anxiety disorders. Some of these biological processes include, inflammation, oxidative and nitrogen stress, and subsequent alteration of neurotrophins, neurogenesis, and neuroplasticity. For instance, PA may promote neuroregeneration, or balance between inflammatory/anti-inflammatory and oxidative/antioxidative marker. This may result to the protective effect against anxiety. Additionally, from a psychological perspective, PA may reduce the risk of developing anxiety through reduced anxiety sensitivity or improved psychological factors such as increased self-efficacy regarding the ability to exert control over potential threats.

The consistent findings from the two review studies included in our qualitative synthesis provide sufficient evidence that an association between PA and anxiety exists. When assessed against Bradford Hill's criteria for causality [45, 46], we graded our findings as probable evidence for a causal relationship. This supports the inclusion of anxiety in the *NSW Active Transport Health Model*.

The main findings of our evidence review for causal relationship are summarised in the Appendix table K-6 below.

Appendix table K-6 Evidence for causal relationship

Criteria	Description
1.Temporality	Both reviews by Schuch and colleagues [50] and McDowell and colleagues the [49] included only prospective cohort studies with a follow up period of more than 1 year. These studies assess prospective evidence where exposure (PA) preceded the effect (anxiety) in time. This evidence meets the temporality criterion: exposure preceded outcome in the included cohort studies.
2. Strength of association	We used measures proposed by Webb and colleagues [45] to guide our classification of the strength of association between PA and anxiety. OR values >0.67 were classified as weak associations, and thus the association (OR 0.74) qualifies as weak. Though the strength of association facilitates assessment for possible causal relationship, a strong association is neither necessary nor sufficient for causality, and weakness is neither necessary nor sufficient for concluding absence of causality.
3. Consistency	In the study by McDowell and colleagues [49] study, all crude and adjusted associations included in the current meta-analyses indicated inverse associations between physical activity and subsequent anxiety. Schuch and colleagues [50] also provide additional evidence of the protective effects of self-reported PA on anxiety development referencing previous cross-sectional studies. We found that there is evidence for repeated observation of an association between PA and anxiety from other studies in different populations under different circumstances. These findings support the criterion of consistency.

4. Dose-response relationship	In their study, Schuch and colleagues [50] did not investigate a dose response relationship between physical activity exposure and anxiety. In the study by McDowell and colleagues [49], a total of 11 included studies assessed for a dose response relationship between PA and various anxiety outcomes. All the 11 studies reported lower odds of anxiety outcomes for increased amounts of PA. In all, there is modest evidence of a dose-response relationship.
5.Biological plausibility	Though the mechanisms are largely unclear, both studies presented evidence of potential biological processes that may underlie the protective effect of PA on incident anxiety.
6. Specificity	As with depression, this criterion is not met but also of questionable relevance.
7.Coherence	The interpretation for the association of PA and anxiety does not conflict with what is known of the natural history and biology of anxiety.
Assessment of grade of evidence Convincing / Probable / Possible / Insufficient	When assessed against Bradford Hill's criteria for causality [45, 46], we considered that the evidence supports a judgement of a *probable* causal relationship. This supports the inclusion of anxiety in the *NSW Active Transport Health Model*.

Appendix L: Physical activity and musculoskeletal disorders

Relationship between PA and health outcomes osteoarthritis and low back pain

Introduction

The aim of this sub-study was to assess whether physical activity improves musculoskeletal health, specifically osteoarthritis and low back pain. We performed a systematic review of systematic reviews, and applied criteria for causal inference to evidence to support a causal interpretation.

Methods

Data Sources and search strategy

The search strategy was informed by guidelines for systematic reviews and rapid systematic reviews [95]. The review is prepared according to the Preferred Reporting Items for Systematic reviews and Meta-Analysis Protocols (PRISMA-P) 2015 statement [96]. The reviews covered the two concept areas: exposure and health outcome. The search was restricted to studies carried out on humans.

The development of the search strategies followed an iterative process. The search was informed by key articles identified in Phase 1 of the project and consultation of a research librarian at Griffith University. The search strategy was tested and modified in PubMed and Embase. We also reviewed reference lists from included studies for suitable studies that met the inclusion criteria. We studied the medical subject heading search (MeSH) terminology in PubMed to inform the selection of our search terms. Appendix table L-1 shows the search strategy that was used.

Appendix table L-1 Search strategies for the three systematic reviews

Exposure and health outcome search terms
("physical activity" OR "physical exercise" OR walking OR bicycling OR "active transport") AND (osteoarthritis OR "lower back pain" OR "low back pain" OR "low backache" OR "back pain" OR "musculoskeletal disease" OR "arthritis" OR "joint pain")

Inclusion and exclusion criteria

The inclusion and exclusion criteria described below were agreed upon by all reviewers.

- Participants/ Population: the review included studies reporting results for whole populations who did not have any of the study outcomes at baseline, covered either all ages, multiple age categories, or specific age groups so long as they are representative of the whole population at that age
- Studies reporting findings of the associations between PA (exposure) and the outcomes: osteoarthritis, low back pain.
- Study designs: systematic reviews and meta-analyses
- Publication status: published studies or studies in print whose full text is publicly available. Most recent versions were considered for studies published in multiple papers.
- Timeline: Studies published in the year 2000 to 28/4/2020
- Language: English
- Studies that provided risk estimates (relative risk, hazard ratio, or odds ratio) with confidence intervals or standard errors, or the data needed to calculate them were included.

Exclusion Criteria

- Based on publication: duplicate publications of the same material were excluded.
- Based on the type of article (e.g., conference abstracts, commentaries, letters)

Study Records

Data management

The identified studies were imported to EndNote X9 software. Duplicate records were identified and excluded. The inclusion and exclusion criteria above guided our study selection. The full text of the selected articles was retrieved and saved into EndNote X9 software.

Screening

One reviewer (MW) screened the titles and abstracts of identified studies for relevance based on the set criteria. One reviewer then screened the full texts of studies that were potentially eligible for inclusion. The other authors reviewed and confirmed the selection process. The final list of studies was discussed and agreed upon by all authors. We documented the reasons for any excluded studies at this (Appendix N).

Data extraction

All reviewers achieved consensus on which data to extract from included studies. One reviewer (MW) extracted data from the full texts of eligible studies. Two other reviewers (HM, LV) cross-checked the data extraction variables. The main areas of data extraction from each study meeting the inclusion criteria are shown in Box 1.

Box 1 Data extraction fields systematic review

- 1st author's last name
- Year of data analysis - dates in search criteria
- Year of publication
- AMSTAR score
- Aim
- Study design selection
- Population
- Reported quality of evidence in the included reviews
- Exposure Details
- Outcome Details
- Data analysis applied
- Effect size reported
- Dose-response results reported
- Publication bias assessed
- Published results of sub-group analyses
- Main conclusion
- Funding
- Other comments

Risk of bias and quality appraisal

We assessed the methodological quality of the included reviews using the Assessment of Multiple Systematic Reviews (AMSTAR) rating scale [97].

Data Synthesis

We identified the most up to date studies that quantified the relationship between physical activity and musculoskeletal diseases (OA and LBP). Study selection was guided by the quality scores for each study, presentation of elaborate details on exposure, outcome, and association measures presented in the papers. Alternative reported values were identified for sensitivity analyses in our model. Our criteria for appraisal of evidence, assessment for causal relationship and grading of evidence is presented in Appendix table L-2.

Appendix table L-2 Criteria for appraisal of evidence, assessment for causal relationship and grading of evidence

Appraising the quality of the evidence presented in the systematic reviews

We used a modified version of GRADE (Grading of Recommendations, Assessment, Development, and Evaluation) guidelines [45].

We applied the 7 *criteria* and '*what to look for'* as described by Webb and colleagues [45] with expanded assessment on our areas of interest such as; measurement of exposure, measurement and definition of outcome, internal validity; random error, chance, by looking at CI ranges, p values, systematic errors, bias by reporting on selection bias, information bias, and confounding. Our appraisals were guided by the reported information in our included studies.

- Criteria
  - xv. Focused Research question
  - xvi. Inclusion and exclusion criteria
  - xvii. Comprehensiveness of search strategy
  - xviii. Assessment of included studies
  - xix. Reproducibility of assessments
  - xx. Similarity of results of include studies
  - xxi. Overall logic and insight

Assessing the reviews against causal criteria

We followed the Bradford Hill criteria [45, 46]. A summarised version of the criteria adopted is outlined below

Criterion	**Description**
1.Temporality	A cause must precede the effect in time. Observation in which cause (exposure) followed effect (outcome) merely shows that the exposure could not have caused the outcome in that instance; it provides no evidence for or against the hypothesis that the exposure can cause the outcome in instances which the exposure precedes the outcome. Only if found that the exposure cannot precede the outcome, can one dispense the causal hypothesis that exposure could cause the outcome.
2. Strength of association	We assessed measures as described by the relative effect, odds ratios or relative risks. We used an adaptation of measures by Webb and colleagues [45] to guide our classification. RR > 3.0 (<0.33), moderately strong; > 5.0 (<0.2), strong. *Additional measures:* RR 1.5-2.9 (0.34-0.67), modest; RR <1.5 (>0.67), weak association. A strong association is neither necessary nor sufficient for causality, weakness is neither necessary nor sufficient for absence of causality.
3. Consistency	Repeated observation of an association from other studies in different populations under different circumstances. Lack of consistency does not rule out a causal association. The effect of a causal agent cannot occur unless the complementary component causes act to complete a sufficient cause. In some circumstances, these conditions may not be met.

4. Dose-response relationships	A dose-response relationship can add weight to an evaluation of causation, but its absence need not count against a causal link. Relationship not always linear.
5.Biological plausibility	If there is a likely biological mechanism through which an exposure might cause the disease. This can add substantial weight to a casual argument. Lack of plausibility does not necessarily rule out causation, because increasing knowledge of disease mechanisms may reveal an association to be more credible with time.
6. Specificity	Where the relationship between exposure and disease is specific. A cause leads to a single effect, not multiple effects, and that an effect has one cause, not multiple causes. This comes into play when it can logically be deduced from the causal hypothesis in question and when non-specificity can be logically deduced from one or more noncausal hypotheses.
7.Coherence	A cause-and-effect interpretation for an association does not conflict with what is known of the natural history and biology of the disease. The absence of coherent information should not be taken as evidence against an association being considered causal. The presence of conflicting information may refute a hypothesis, but the conflicting information may be mistaken or misinterpreted.

Grading the evidence

We graded the evidence to support a judgement of a relationship. This grading process was guided by the World Cancer Research Fund grading system [23] with modification to best align to the nature of evidence that we assessed in our systematic review of reviews. A modification of the WCRF grading system was applied in the GBD 2017 study [22]. We adopted grades of convincing, probable, possible and insufficient evidence.

Convincing (strong) evidence

Evidence strong enough to support a judgement of a convincing causal (or protective) relationship, which justifies making recommendations designed to reduce the risk of health outcome. All the following are generally required:

- Evidence from at least two independent cohort studies
- No substantial unexplained heterogeneity within or between study types or in different populations relating to the presence or absence of an association, or direction of effect.
- Good quality studies to exclude with confidence the possibility that the observed association results from random or systematic error, inducing confounding, measurement error and selection bias.
- Evidence showing consistent associations between exposure and disease, with little or no evidence to the contrary
- The association should be biologically plausible.
- Presence of a plausible biological gradient ('dose-response') in the association. Such a gradient need not be linear or even in the same direction across the different levels of exposure, so long as this can be explained plausibly.

Probable evidence

Evidence strong enough to support a judgement of a probable causal (or protective) relationship, which generally justifies recommendations designed to reduce the risk of health outcome.

Evidence based on epidemiological studies showing fairly consistent associations between exposure and disease, but for which there are perceived shortcomings in the available evidence or some evidence to the contrary, which precludes a more definite judgment. Shortcomings in the evidence may be any of the following: insufficient duration of or studies; insufficient or studies available; inadequate sample sizes; or incomplete follow-up.

All the following are generally required:

- Evidence from at least two independent cohort studies
- No substantial unexplained heterogeneity within or between study types or in different populations relating to the presence or absence of an association, or direction of effect.
- Good quality studies to exclude with confidence the possibility that the observed association results from random or systematic error, inducing confounding, measurement error and selection bias.
- The association should be biologically plausible.
- Presence of a plausible biological gradient ('dose-response') in the association. Such a gradient need not be linear or even in the same direction across the different levels of exposure, so long as this can be explained plausibly.

Possible (suggestive) evidence

Evidence that is too limited to permit a probable or convincing causal judgement but is suggestive of a direction of effect. The evidence may be limited in amount or by methodological flaws but shows a generally consistent direction of effect. This judgement is broad and includes associations whether the evidence falls only slightly below that required to infer a probably causal association through to those where the evidence is only marginally strong enough to identify a direction of effect. This judgement is very rarely sufficient to justify recommendations designed to reduce the risk of health outcome; any exception to this require special, explicit justification.

All the following are generally required:

- Evidence from at least two independent cohort studies
- The direction of effect is generally consistent though some unexplained heterogeneity may be present
- Evidence for biological plausibility

Insufficient evidence

Evidence based on findings of a few studies which are suggestive, but insufficient to establish an association between exposure and disease. More well-designed research is needed to support the tentative association.

Evidence is so limited that no firm conclusion can be made. This judgement represents an entry level and is intended to allow any exposure for which there are sufficient data to warrant consideration, but where insufficient evidence exists to permit a more definitive grading. The

evidence may be limited by the amount of evidence in terms of the number of studies available, by inconsistency of effect, by methodological flaws (for example, lack of adjustment for known confounders) or by any combination of these factors.

This grading does not necessarily mean a judgement that there is evidence of no relationship. With further good quality research, any exposure graded in this way might in the future be shown to increase or decrease the risk of the health outcome under investigation.

Results: Physical activity and health outcomes osteoarthritis and low back pain

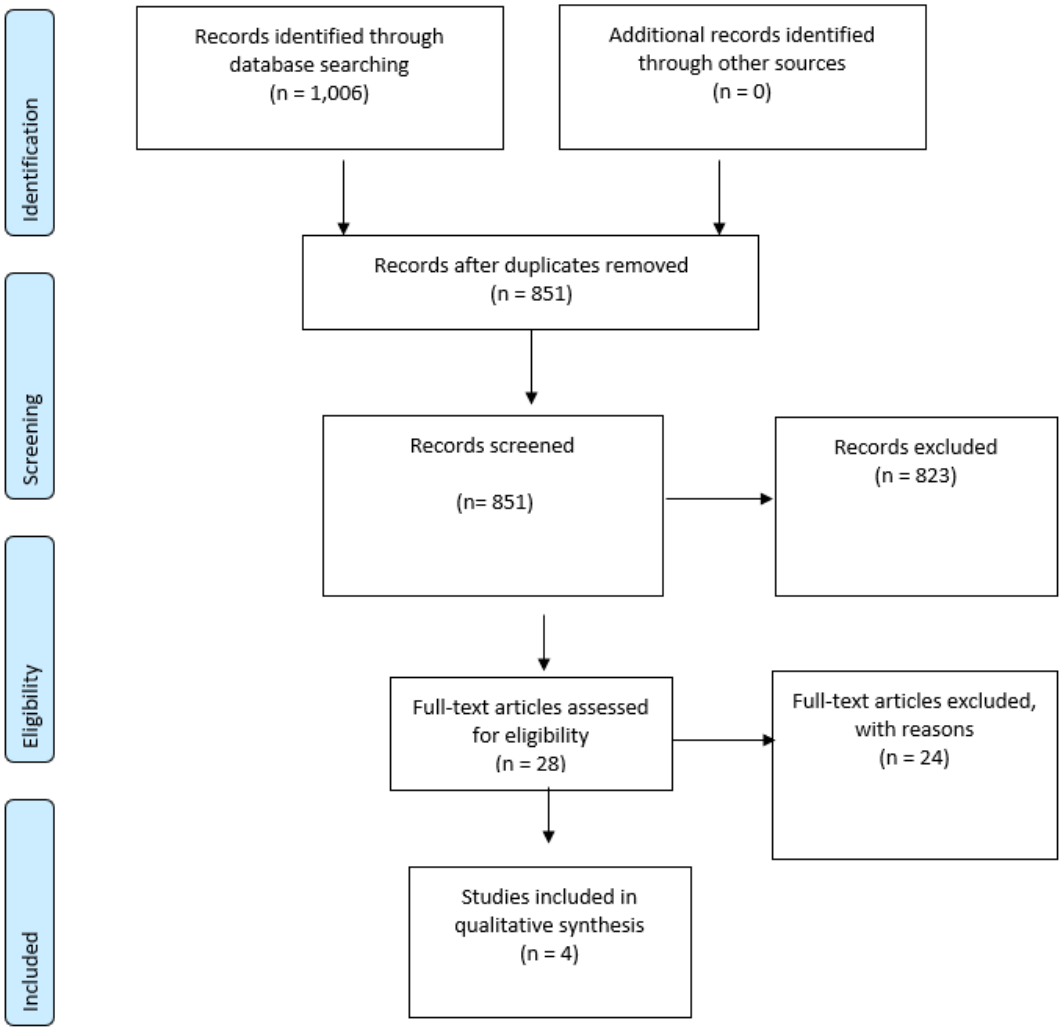
Literature database search

The total number of articles identified is presented in the PRISMA [96] diagram below (Appendix Figure L-1). The database search identified 1,006 articles. A total of 851 records remained after removing duplicates. After screening of title and abstract, 823 records were excluded and 28 remained for full text analysis. A further 23 studies were excluded with reason after full text analysis, leaving a total of 5 studies [51-55] for data extraction.

Reasons for exclusion were that studies did not meet the inclusion criteria for population (n=2), the study design did not meet our inclusion criterion (n=14), the study did not provide measures of disease association describing the relationship between our exposure and study outcome (n=4), the outcome criterion was not met (n=2), or the study exposure variable was not PA (n=1). On closer inspection the Heneweer, Staes [53] study was excluded from our qualitative analysis. This is because the authors presented a summary of the available evidence across studies without giving combined measures of associations for the relationship between PA and LBP. Also, we found that although the authors applied a systematic approach in selection of studies, majority of the retrieved studies focused on occupational workload PA.

Appendix N provides a list of the studies excluded after full-text analysis and specific reasons for exclusion. Using the AMSTAR criteria [97], the two included studies that reported on the outcome low back pain were scored as high quality reviews [51, 52]. while the two studies that reported on OA, received a moderate quality score [54, 55]. Appendix O summarizes the quality scores for each of the included studies.

The aim of this sub-study was to assess whether physical activity reduces the risk for all-cause mortality. We performed a systematic review of systematic reviews, and applied criteria for causal inference to evidence to support a causal interpretation.



Appendix Figure L-1 Prisma flow diagram Physical activity and health outcomes osteoarthritis and low back pain

Low Back Pain

Strength of association

Two systematic review studies [51-53] reported evidence suggestive of an association between PA and LBP. Nevertheless, there were some mixed results across various variables investigated in the two studies. In a meta-analysis of seven cohort studies, Alzahrani and colleagues [51] found that compared to low level total PA, medium level total PA was significantly associated with a decreased risk of developing LBP (RR = 0.90, 95% CI 0.85 to 0.96). However, in a meta-analysis of nine cohort studies, compared to low level total PA, high level total PA was not associated with LBP (pooled risk ratio 1.00, 95% CI 0.92 to 1.08). Shiri and Falah-Hassani [52] conducted a systematic review and meta-analysis of 36 prospective cohort studies to assess the effect of leisure time physical activity (LTPA) on non-specific LBP. Their results indicated that LTPA was neither associated with LBP in the past month nor associated with LBP in the past 6–12 months. However, they found that moderately and highly active individuals had a reduced risk for frequent or chronic LBP when compared against individuals with no regular PA (Appendix table L-3). A summary of the measures of disease association found in our review study is presented in Appendix table L-. The adjusted risk ratios describing the association between PA and frequent or LBP in the Shiri and Falah-Hassani [52] study were selected for use in the model.

Appendix table L-3 Summary of the measures of disease association

Study	**Findings**
Low back pain	
Alzahrani, et al., [51]	• Compared to low level total physical activity, medium level total physical activity[a] was significantly associated with a decreased risk of developing low back pain. Fully adjusted risk ratios, 0.90, 95% CI 0.85 to 0.96 (n=7) [*]. • Compared to low level total physical activity, high level total physical activity[b] was not associated with low back pain. Pooled 1.00, 95% CI 0.92 to 1.08 (n=9) [*]. •
Shiri & Falah-Hassani, 2017 [52]	• Leisure time physical activity (LTPA) was neither associated with low back pain in the past month nor associated with low back pain in the past 6–12 months LTPA and Low back pain in the past month Active[c] Vs inactive (n=5) RR 0.98 95% CI 0.84 to 1.14 Moderate[d] Vs low activity (n=3) RR 0.93 95% CI 0.80 to 1.08 High[e] Vs Low activity (n=4) RR 0.85 95% CI 0.53 to 1.37 LTPA and Low back pain in the past 6–12 months Active[c] Vs inactive (n=16) RR 0.97, 95% CI 0.92 to 1.01 Moderate[d] Vs low activity (n=10) RR 0.99, 95% CI 0.87 to 1.07 High[e] Vs Low activity (n=11) RR 0.90, 95% CI 0.77 to 1.05 High[e] Vs Low or moderate[d]activity (n= 5) RR 1.01, 95% CI 0.89 to 1.15 •

	• LTPA was associated with frequent or chronic low back pain as per adjusted risk ratios below;
Active[c] Vs inactive (n=9)	RR 0.84, 95% CI 0.77 to 0.92
Moderate[d] Vs low activity (n=3)	RR 0.86, 95% CI 0.79 to 0.94
High[e] Vs Low activity (n=3)	RR 0.70, 95% CI 0.48 to 1.03
	• Adjusted RRs for associations between physical activity and frequent or chronic low back pain (Ref group: individuals with no regular physical activity)
physically active individuals	RR=0.89, 95% CI 0.82 to 0.97 (n=6)
moderately active individuals	RR=0.86, 95% CI 0.79 to 0.94 (n=2)
highly active individuals	RR=0.84, 95% CI 0.75 to 0.93 (n=2)

[a] Between the 33th and 66th percentile (reporting >11.89 and <21 MET-hours/week).

[b] Greater than or equal to the 66th percentile (reporting ≥21 MET-hours/week)

[*] Results from included cohort studies

[c] being physically active was defined as participation in a sport or other physical activity during leisure time, at least 1–2 times a week, at least 0.5–1.0 hour per week, or being in the middle or upper third of the distribution of leisure time physical activity in a study sample.

[d] Moderate level of physical activity was defined as participation in such activity 1–3 times a week,..1–3 hours per week..or being in the middle third of the distribution of leisure time physical activity in a study sample.

[e]High level of physical activity was defined as participation in leisure time physical activity, ≥ 3–4 times per week,..more than 2–4 hours per week...or being in the upper third of the distribution of such activity in a study sample

Abbreviations: RR, risk ratios; CI, confidence interval

Appraisal of evidence and assessment of a causal relationship

Two studies [51, 52] investigating the association between physical activity and low back pain met the inclusion criteria for our qualitative synthesis.

In a systematic review and metanalysis of 35 observational studies (cohort and cross-sectional studies), Alzahrani and colleagues [51] investigated the association between total and domain-specific physical activity and non-specific low back pain in adults. A total of 24 high quality studies were included in the quantitative syntheses (15 cohort and 9 cross-sectional). For this review of reviews, only results from the cohort studies were considered. This was to enable assessment of a possible causal relationship that is best done with longitudinal prospective evidence.

Only two studies used objective measurements of physical activity. This is likely to have produced recall bias and overestimation of PA exposure. Also, measurements and classifications of physical activity in terms of frequency, intensity, and duration differed across included studies. This may have led to misclassification of PA levels. To facilitate the integration of activities differing in intensity and duration, the authors harmonized physical activity variables to the common unit of metabolic equivalent (MET)-hours/week. For some studies that did not report the intensity, duration, or frequency for the measured PA, the authors followed standard rules for assigning the dose in MET-hours/week for all PA variables extracted from the included studies.

In a meta-analysis of seven cohort studies, the authors found that, when compared to low level total PA, medium level PA was significantly associated with a decreased risk of developing LBP (RR = 0.90, 95% CI 0.85 to 0.96). High level physical activity was not associated with LBP in a meta-analysis of nine cohort studies (RR = 1.00, 95% CI 0.92 to 1.08). For medium level versus low level leisure-time physical activity, a meta-analysis of six cohort studies showed that medium level LTPA was inversely associated with LBP (RR = 0.90, 95% CI 0.85 to 0.96). For high level versus low level leisure-time physical activity, the meta-analysis of cohort studies did not find an association between high level LTPA and LBP (RR = 1.01, 95% CI 0.93 to 1.10). The authors suggested that the absence of an association between high level LTPA and LBP might be due to including different types and durations of LTPA which may have resulted in misclassification, and then underestimation of the effect of high level LTPA. Authors in this study found no evidence of dose-response relationship.

Only fully adjusted models from each study were included in the analyses to reduce the possibility of confounding. However, the adjusted models varied across the included studies, and the authors were not able to adjust for some potential risk factors, such as occupational and non-occupational physical activity types. Most of these studies were adjusted for age (n = 20), gender (n = 19), and smoking (n = 14). In less than 50% of studies, other potential confounding factors were considered. These factors were body mass index (BMI), education, ethnicity, income, stress, anxiety, depression, fear of pain, alcohol consumption, occupation, self-rated health, overweight, obesity, sleep quality, chronic disease history, cardiorespiratory, hypercholesterolemia, diabetes, hypertension, pain management, medication use, consultation, musculoskeletal symptoms or injuries, surgery, disability, total hip osteoporosis status, residential area, nutritional level, fitness, father's occupation, whole body vibration, occupational activities, and other types of activities.

The authors identified the following potential sources of bias; nine cohort studies (out of 15) failed to describe the characteristics of patients who were lost to follow-up, and eleven cohort studies did not use a reliable measure of physical activity outcome which may impact the internal validity. Additionally, seven cohort studies failed to report the proportion of participants who agreed to participate in the study and whether they were representative of the source population which may impact on generalisability of the study results. However, when the authors restricted the analyses to

very high-quality studies, the results of the meta-analyses did not change suggesting the results were robust against bias. Alzahrani and colleagues [51] concluded that their results provided evidence suggesting that there is an inverse association between physical activity and LBP.

In the second study, Shiri and Falah-Hassani [52] conducted a systematic review and meta-analysis of 36 prospective cohort studies to assess the effect of leisure time physical activity (LTPA) on non-specific LBP. For the outcome LBP, the authors investigated LBP in the past month or the past 6–12 months, and frequent or chronic LBP. Chronic LBP was defined as pain that had lasted for 3 months or longer or pain for more than 30 days in the past 12 months. In the metanalysis, the studies on chronic LBP, except one, included participants free from chronic LBP at baseline. This limited the possibility of reverse causation bias.

Quality assessment for the included studies was done using criteria adapted from the Effective Public Health Practice Project tool. The authors assessed four sources of bias: selection bias, performance bias, attrition bias, and confounding. For selection bias, the ratings were as follows; low risk (n=8), moderate risk (n=19), and high risk (n=9). Nine studies were assessed as having a low risk of performance bias, 23 were assessed as having moderate risk and four were assessed as having a high risk of performance bias. Fifteen studies were rated as having a low risk of attrition bias, 14 were rated as having moderate risk and seven were rated as having high risk. Twenty-eight studies controlled their risk estimates for some confounding factors such as age, sex, and body mass index.

The measurement and classification of PA varied across the included studies. Only two studies measured physical activity by an accelerometer, and only three studies defined the levels of physical activity by using a physical activity index or metabolic equivalent of task (MET). Different cut points for moderate and high levels of LTPA were used. For instance, the authors point out that several studies classified individuals with 2-3 hours per week LTPA as inactive. This exposure misclassification may have led to the underestimation of an association between LTPA and LBP.

The authors found that LTPA was not associated with changes in risk of LBP in the past month or the past 6–12 months. However, participation in a sport or other leisure time physical activity was inversely associated with frequent or chronic LBP. From the six studies included in the meta-analysis, the risk of frequent or chronic LBP was lower by 11% (RR=0.89, CI 0.82 to 0.97) in physically active individuals, 14% (RR=0.86, CI 0.79 to 0.94) in moderately active individuals and by 16% (RR=0.84, CI 0.75 to 0.93) in highly active individuals, all in comparison with individuals with no regular physical activity. These six studies controlled the risk estimates for the following potential confounders; age (n=6), sex and body mass index (n=5), smoking (n=2), educational level (n=1), job demands (n=1), social class (n=1), and occupation(n=1). Notably, among all the included studies (n=36), those that did not adjust their risk estimates for any confounding factors reported a stronger inverse association between physical activity and frequent or chronic LBP than those studies that controlled their estimates for some confounders. Not adjusting for known confounders may have led to an overestimation of the protective effect of leisure physical activity on frequent or chronic LBP. For the studies included in the metanalysis, sensitivity analyses were performed with regard to the presence or absence of LBP at baseline, age of participants, adjustment for confounding factors, and other methodological quality of included studies. The sensitivity analysis showed similar associations in adults and elderly people. For the rest of the confounding factors, the pooled estimates changed only slightly after excluding the studies that reported unadjusted risk estimates. Overall, the level of heterogeneity was low to moderate. The authors record that heterogeneity across studies on frequent or chronic LBP was explained fully by selection bias and adjustment for confounding factors. Heterogeneity across studies on LBP in the past 1–12 months was partly explained by selection bias. No publication bias was seen for LBP in the past 1–12 months as well as for frequent or chronic LBP.

The authors considered their findings to suggest that moderate to high level of physical activity during leisure time protects against frequent or chronic LBP [52].

From the consistent findings in the two review studies included in our qualitative synthesis [51, 52], there was sufficient evidence indicative of an association between PA and LBP. When assessed against Bradford Hill's criteria for causality we graded our findings as probable evidence for a causal relationship. This supports the inclusion of LBP in the *NSW Active Transport Health Model*. A summary of the assessment for evidence for a causal relationship is given below Appendix table L-4.

Appendix table L-4 Assessing the evidence against causal criteria: Low back pain

Criteria	Description
1.Temporality	From the two systematic review and meta-analyses studies included in our study [51, 52], we appraised evidence from prospective cohort studies. Inclusion of prospective studies limited the possibility of reverse causation bias allowing for the examination of the temporal sequence between baseline levels of PA and LBP. We considered the evidence from these studies to have met the temporality criteria where the exposure (PA) was seen to precede the effect (LBP).
2. Strength of association	The results for the associations specific associations found between PA and LBP are reported in detail in Appendix 9. The findings provide evidence that suggests that there is an inverse association between PA and LBP. For instance in the study by Shiri & Falah-Hassani [52], the risk of frequent/chronic LBP was 14% lower (RR=0.86, CI 0.79 to 0.94, I2=0%, n=33 032) in moderately active individuals and 16% lower (RR=0.84, CI 0.75 to 0.93, I2=0%, n=33 032) highly active individuals in comparison with individuals without regular physical activity. Using an adaptation of measures by Webb and colleagues [45] to guide our classification of the strength of association, we considered the reported measures of strength of association weak. Weakness of an association makes the risk of alternative explanations greater but does not preclude causality.
3. Consistency	Our findings provide cautious support for the criterion of consistency. Alzahrani and colleagues [51] concluded that their results provided evidence suggesting that there is an inverse association between physical activity and LBP. Compared to low level total physical activity, medium level total physical activity was significantly associated with a decreased risk of developing low back pain (fully adjusted risk ratios, 0.90, 95% CI 0.85 to 0.96). Shiri and Falah-Hassani [52] considered their findings to suggest that moderate to high level of physical activity during leisure time protects against frequent or chronic LBP.
4. Dose-response relationship	In their systematic review and meta-analysis, Alzahrani and colleagues [51] found no evidence of dose-response relationship. Shiri and Falah-Hassani [52] did not investigate a dose response relationship.
5.Biological plausibility	There are several proposed biological mechanisms by which PA is associated with LBP. Though strenuous PA has been considered to increase the risk of LBP, the kind of PA involved in active transport is unlikely to have the same impact. Instead this type of PA may lead to

	increased muscle strength and flexibility hence protecting the spine from injuries.
6. Specificity	Not supported but not highly applicable.
7.Coherence	The interpretation for the association of PA and LBP does not conflict with what is known of the natural history and biology of LBP.
Assessment of grade of evidence Convincing / Probable / Possible / Insufficient	From the consistent findings in the two review studies included in our qualitative synthesis [51, 52], there was sufficient evidence indicative of an association between PA and LBP. When assessed against Bradford Hill's criteria for causality we graded our findings as possible (suggestive) evidence for a causal relationship. This supports the inclusion of LBP in the *NSW Active Transport Health Model* but it should be reserved for sensitivity analyses and not be included in the main analysis.

As discussed earlier, the two studies included for the outcome OA [54, 55] did not yield combined measures of association. We, therefore, carried out a new review of the literature using a modified protocol to identify cohort studies investigating the association between walking, cycling, or 'active transport' and OA. Appendix table L-5 summarises the search process applied.

Appendix table L-5 The search process to identify cohort studies investigating the association between active transport and osteoarthritis

Database	Search terms	Restrictions applied	**Results** (8/5/2020)
Pub Med	(walking OR bicycling OR "active transport") AND (osteoarthritis OR "musculoskeletal disease" OR "arthritis" OR "joint pain")	• Article type: Observational study (best option available for the cohort study restriction). • Publication dates: 2000 to date • Language: English • Search field: Title/Abstract	45
Embase	(walking OR bicycling OR "active transport") AND (osteoarthritis OR "musculoskeletal disease" OR "arthritis" OR "joint pain")	• Restrictions: • Search filed: Title/Abstract • Study types: cohort analysis, prospective study, observational study, • Publication type- article, article in press • Language: English	57

The database search identified 102 studies (Appendix table L-5). A total of 93 records remained after removing 9 duplicates. After screening of title and abstract, 90 records were excluded and 3 remained for full text analysis. On closer inspection of the full text of the 3 studies, a further 2 studies [98, 99] were excluded with reason leaving 1 study [56] for data extraction. The included study was a

prospective cohort study by White and colleagues [56] that examined the association of step-defined daily walking with incident functional limitation two years later in people with or at risk of knee OA.

Appraisal of evidence and assessment of causal relationship – OA

White and colleagues [56] examined the association of step-defined daily walking with incident functional limitation two years later in people with or at risk of knee OA . The authors pooled data from 6 prospective cohort studies. The median follow up time was 14.2 years. Only study participants without functional limitation at baseline were included. In our systematic review of reviews study, we assumed that the risk of functional limitation equals the risk of developing OA. Incident functional limitation over two years was defined by performance-based (gait speed 1.0 m/s) and self-report measures. Gait speed was measured as the average of two trials of walking along a marked 20-meter course in an unobstructed corridor. Walking over 7 days at baseline was objectively measured as steps/day using a StepWatch Activity Monitor; a waterproof, self-contained accelerometer-based device. To ensure a reliable estimate of PA, the authors restricted their sample to those participants who wore the device for at least 3 valid days.

The authors in this study found that a greater number of steps/days, measured either by self-report or performance based, was increasingly protective against the development of functional limitation. For our review study, we utilized the results for the performance based definition of functional limitation by gait speed. People who walked 5000 – 7499 steps/day by performance-based measure had 0.50 times the adjusted risk compared with those walking less than 5000 steps/day. Those who walked 7500 steps/day, had 0.31 times the risk, compared with those walking less than 5000 steps/day. Their findings indicated a dose response relationship between PA and functional limitation. For incident functional limitation by the performance-based, each additional 1000 steps/day was associated with a 16% reduction [56]. Potential confounders that the authors considered were; age, sex, race, education, body mass index (BMI), Radiographic knee OA (ROA), knee pain, self-reported comorbidities, depressive symptoms and widespread pain.

The average steps/day values used to estimate pedometer step counts in this study were similar to previously published studies. Nonetheless, the authors proposed careful interpretation of their results since different monitors and methods of estimating pedometer steps have been previously employed across studies. Since this was a cohort study with fixed time-point visit, the outcome of incident functional limitation did not discriminate between those who truly had the first instance of functional limitation at two years from those who had fluctuating physical function over time. To address this, the authors followed the standard approach used to assess this type of outcome. In this study, White and colleagues [56] note that there may have been the possibility that participants engaged in PA with no steps such as cycling, swimming or minimal steps such as gardening. Nonetheless, this was considered unlikely basing it on the results of data previously collected from study participants where 9% of men and 7% of women had reported engaging in strenuous non-ambulatory activities ‘often’.

The findings of this large cohort study were consistent with previous studies discussed by the authors. They also noted that the PA thresholds for functional limitation found in the study were consistent with the findings for other clinical outcomes such as adverse cardiometabolic health indicators. Appendix table L-6 gives a summary of the assessed evidence against causal criteria.

Appendix table L-6 Assessing the evidence against causal criteria: Osteoarthritis

Criteria	Description
1.Temporality	In the large cohort study by White and colleagues [56], data was pooled data from 6 prospective cohort with a median follow up time of 14.2 years. Only study participants without functional limitation at baseline were included. Inclusion of prospective studies limited the possibility of reverse causation bias allowing for the examination of the temporal sequence between PA and functional limitation. Since OA is usually diagnosed based on both functional limitations and imaging (X-rays or MRI), functional limitation was taken to represent OA in our study. We considered the evidence to have met the temporality criteria where the exposure (PA) was seen to precede the effect (OA).
2. Strength of association	The findings by White and colleagues [56] indicated that a greater number of steps/days, measured either by self-report or performance based, was increasingly protective against the development of functional limitation. Using an adaptation of measures by Webb and colleagues [45] to guide our classification of the strength of association, the reported measures of strength of association ranged from modest to moderately strong. Weakness of an association makes the risk of alternative explanations greater but does not preclude causality.
3. Consistency	The findings were consistent with previous studies discussed by the authors. The authors also highlight that the PA thresholds for functional limitation found in the study were consistent with the findings for other clinical outcomes such as adverse cardiometabolic health indicators.
4. Dose-response relationship	The findings give evidence of a dose response relationship between PA and functional limitation
5.Biological plausibility	In our study, we assumed that the risk of functional limitation equals the risk of developing OA. Repeated intense stress on the joints, or acute high stress, which can occur in manual labour or intense sport training, may damage cartilage, ligaments and other joint structures, which may lead to OA. In contrast, the kind of PA involved in active transport is unlikely to have the same impact, but instead stabilise the joints by strengthening muscles.
6. Specificity	Not supported, not readily applicable.
7.Coherence	The interpretation for the association of PA and OA does not conflict with what is known of the natural history and biology of OA.
Assessment of grade of evidence Convincing / Probable / Possible / Insufficient	We graded our findings as possible (suggestive) evidence for a causal relationship. This supports the inclusion of OA in the *NSW Active Transport Health Model*. However, the earlier assumption made on the outcome, functional limitation being equal to OA, is a limitation to our study. We therefore propose to include the OA only in the sensitivity analysis for the model.

Appendix M: Physical activity and all-cause mortality

Introduction

The aim of this sub-study was to assess whether physical activity reduces the risk for all-cause mortality. We performed a systematic review of systematic reviews, and applied criteria for causal inference to evidence to support a causal interpretation.

Methods

Data Sources and search strategy

The search strategy was informed by guidelines for systematic reviews and rapid systematic reviews [95]. The review is prepared according to the Preferred Reporting Items for Systematic reviews and Meta-Analysis Protocols (PRISMA-P) 2015 statement [96]. The reviews covered the two concept areas: exposure and health outcome. The search was restricted to studies carried out on humans.

The development of the search strategies followed an iterative process. The search was informed by key articles identified in Phase 1 of the project and consultation of a research librarian at Griffith University. The search strategy was tested and modified in PubMed and Embase. We also reviewed reference lists from included studies for suitable studies that met the inclusion criteria. We studied the medical subject heading search (MeSH) terminology in PubMed to inform the selection of our search terms. Appendix table M-1 shows the search strategy that was used.

Appendix table M-1 Search strategies for the three systematic reviews

Exposure and health outcome search terms
(“physical activity” OR “physical exercise” OR walking OR bicycling OR “active transport”) AND (“all-cause mortality” OR mortality OR deaths)

Inclusion and exclusion criteria

The inclusion and exclusion criteria described below were agreed upon by all reviewers.

- Participants/ Population: the review included studies reporting results for whole populations who did not have any of the study outcomes at baseline, covered either all ages, multiple age categories, or specific age groups so long as they are representative of the whole population at that age
- Studies reporting findings of the associations between PA (exposure) and the outcome, ACM.
- Study designs: systematic reviews and meta-analyses
- Publication status: published studies or studies in print whose full text is publicly available. Most recent versions were considered for studies published in multiple papers.
- Timeline: Studies published in the year 2000 to 19/3/2020
- Language: English
- Studies that provided risk estimates (relative risk, hazard ratio, or odds ratio) with confidence intervals or standard errors, or the data needed to calculate them were included.

Exclusion Criteria

- Based on publication: duplicate publications of the same material were excluded.
- Based on the type of article (e.g., conference abstracts, commentaries, letters)

Study Records

Data management

The identified studies were imported to EndNote X9 software. Duplicate records were identified and excluded. The inclusion and exclusion criteria above guided our study selection. The full text of the selected articles was retrieved and saved into EndNote X9 software.

Screening

One reviewer (MW) screened the titles and abstracts of identified studies for relevance based on the set criteria. One reviewer then screened the full texts of studies that were potentially eligible for inclusion. The other authors reviewed and confirmed the selection process. The final list of studies was discussed and agreed upon by all authors. We documented the reasons for any excluded studies at this stage (Appendix N).

Data extraction

All reviewers achieved consensus on which data to extract from included studies. One reviewer (MW) extracted data from the full texts of eligible studies. Two other reviewers (HM and LV) cross-checked the data extraction variables. The main areas of data extraction from each study meeting the inclusion criteria are shown in Box 1.

Box 1 Data extraction fields systematic review

- 1st author's last name
- Year of data analysis - dates in search criteria
- Year of publication
- AMSTAR score
- Aim
- Study design selection
- Population
- Reported quality of evidence in the included reviews
- Exposure Details
- Outcome Details
- Data analysis applied
- Effect size reported
- Dose-response results reported
- Publication bias assessed
- Published results of sub-group analyses
- Main conclusion
- Funding
- Other comments

Risk of bias and quality appraisal

We assessed the methodological quality of the included reviews using the Assessment of Multiple Systematic Reviews (AMSTAR) rating scale [97].

Data Synthesis

We identified the most up to date studies that quantified the relationship between PA and ACM. Study selection was guided by the quality scores for each study, presentation of elaborate details on exposure, outcome, and association measures presented in the papers. Alternative reported values were identified for sensitivity analyses in our model. Our criteria for appraisal of evidence, assessment for causal relationship and grading of evidence is presented in Appendix table M-2.

Appendix table M-2 Criteria for appraisal of evidence, assessment for causal relationship and grading of evidence

Appraising the quality of the evidence presented in the systematic reviews

We used a modified version of GRADE (Grading of Recommendations, Assessment, Development, and Evaluation) guidelines [45].

We applied the 7 *criteria* and '*what to look for'* as described by Webb and colleagues [45] with expanded assessment on our areas of interest such as; measurement of exposure, measurement and definition of outcome, internal validity; random error, chance, by looking at CI ranges, p values, systematic errors, bias by reporting on selection bias, information bias, and confounding. Our appraisals were guided by the reported information in our included studies.

- Criteria
  - xxii. Focused Research question
  - xxiii. Inclusion and exclusion criteria
  - xxiv. Comprehensiveness of search strategy
  - xxv. Assessment of included studies
  - xxvi. Reproducibility of assessments
  - xxvii. Similarity of results of include studies
  - xxviii. Overall logic and insight

Assessing the reviews against causal criteria

We followed the Bradford Hill criteria [45, 46]. A summarised version of the criteria adopted is outlined below

Criterion	Description
1.Temporality	A cause must precede the effect in time. Observation in which cause (exposure) followed effect (outcome) merely shows that the exposure could not have caused the outcome in that instance; it provides no evidence for or against the hypothesis that the exposure can cause the outcome in instances which the exposure precedes the outcome. Only if found that the exposure cannot precede the outcome, can one dispense the causal hypothesis that exposure could cause the outcome.

2. Strength of association	We assessed measures as described by the relative effect, odds ratios or relative risks. We used an adaptation of measures by Webb and colleagues [45] to guide our classification. RR > 3.0 (<0.33), moderately strong; > 5.0 (<0.2), strong. *Additional measures:* RR 1.5-2.9 (0.34-0.67), modest; RR <1.5 (>0.67), weak association. A strong association is neither necessary nor sufficient for causality, weakness is neither necessary nor sufficient for absence of causality.
3. Consistency	Repeated observation of an association from other studies in different populations under different circumstances. Lack of consistency does not rule out a causal association. The effect of a causal agent cannot occur unless the complementary component causes act to complete a sufficient cause. In some circumstances, these conditions may not be met.
4. Dose-response relationships	A dose-response relationship can add weight to an evaluation of causation, but its absence need not count against a causal link. Relationship not always linear.
5.Biological plausibility	If there is a likely biological mechanism through which an exposure might cause the disease. This can add substantial weight to a casual argument. Lack of plausibility does not necessarily rule out causation, because increasing knowledge of disease mechanisms may reveal an association to be more credible with time.
6. Specificity	Where the relationship between exposure and disease is specific. A cause leads to a single effect, not multiple effects, and that an effect has one cause, not multiple causes. This comes into play when it can logically be deduced from the causal hypothesis in question and when non-specificity can be logically deduced from one or more noncausal hypotheses.
7.Coherence	A cause-and-effect interpretation for an association does not conflict with what is known of the natural history and biology of the disease. The absence of coherent information should not be taken as evidence against an association being considered causal. The presence of conflicting information may refute a hypothesis, but the conflicting information may be mistaken or misinterpreted.

Grading the evidence

We graded the evidence to support a judgement of a relationship. This grading process was guided by the World Cancer Research Fund grading system [23] with modification to best align to the nature of evidence that we assessed in our systematic review of reviews. A modification of the WCRF grading system was applied in the GBD 2017 study [22]. We adopted grades of convincing, probable, possible and insufficient evidence.

Convincing (strong) evidence

Evidence strong enough to support a judgement of a convincing causal (or protective) relationship, which justifies making recommendations designed to reduce the risk of health outcome. All the following are generally required:

- Evidence from at least two independent cohort studies
- No substantial unexplained heterogeneity within or between study types or in different populations relating to the presence or absence of an association, or direction of effect.
- Good quality studies to exclude with confidence the possibility that the observed association results from random or systematic error, inducing confounding, measurement error and selection bias.
- Evidence showing consistent associations between exposure and disease, with little or no evidence to the contrary
- The association should be biologically plausible.
- Presence of a plausible biological gradient ('dose-response') in the association. Such a gradient need not be linear or even in the same direction across the different levels of exposure, so long as this can be explained plausibly.

Probable evidence

Evidence strong enough to support a judgement of a probable causal (or protective) relationship, which generally justifies recommendations designed to reduce the risk of health outcome.

Evidence based on epidemiological studies showing fairly consistent associations between exposure and disease, but for which there are perceived shortcomings in the available evidence or some evidence to the contrary, which precludes a more definite judgment. Shortcomings in the evidence may be any of the following: insufficient duration of or studies; insufficient or studies available; inadequate sample sizes; or incomplete follow-up.

All the following are generally required:

- Evidence from at least two independent cohort studies
- No substantial unexplained heterogeneity within or between study types or in different populations relating to the presence or absence of an association, or direction of effect.
- Good quality studies to exclude with confidence the possibility that the observed association results from random or systematic error, inducing confounding, measurement error and selection bias.
- The association should be biologically plausible.
- Presence of a plausible biological gradient ('dose-response') in the association. Such a gradient need not be linear or even in the same direction across the different levels of exposure, so long as this can be explained plausibly.

Possible (suggestive) evidence

Evidence that is too limited to permit a probable or convincing causal judgement but is suggestive of a direction of effect. The evidence may be limited in amount or by methodological flaws but shows a generally consistent direction of effect. This judgement is broad and includes associations whether the evidence falls only slightly below that required to infer a probably causal association through to those where the evidence is only marginally strong enough to identify a direction of

effect. This judgement is very rarely sufficient to justify recommendations designed to reduce the risk of health outcome; any exception to this require special, explicit justification.

All the following are generally required:

- Evidence from at least two independent cohort studies
- The direction of effect is generally consistent though some unexplained heterogeneity may be present
- Evidence for biological plausibility

Insufficient evidence

Evidence based on findings of a few studies which are suggestive, but insufficient to establish an association between exposure and disease. More well-designed research is needed to support the tentative association.

Evidence is so limited that no firm conclusion can be made. This judgement represents an entry level and is intended to allow any exposure for which there are sufficient data to warrant consideration, but where insufficient evidence exists to permit a more definitive grading. The evidence may be limited by the amount of evidence in terms of the number of studies available, by inconsistency of effect, by methodological flaws (for example, lack of adjustment for known confounders) or by any combination of these factors.

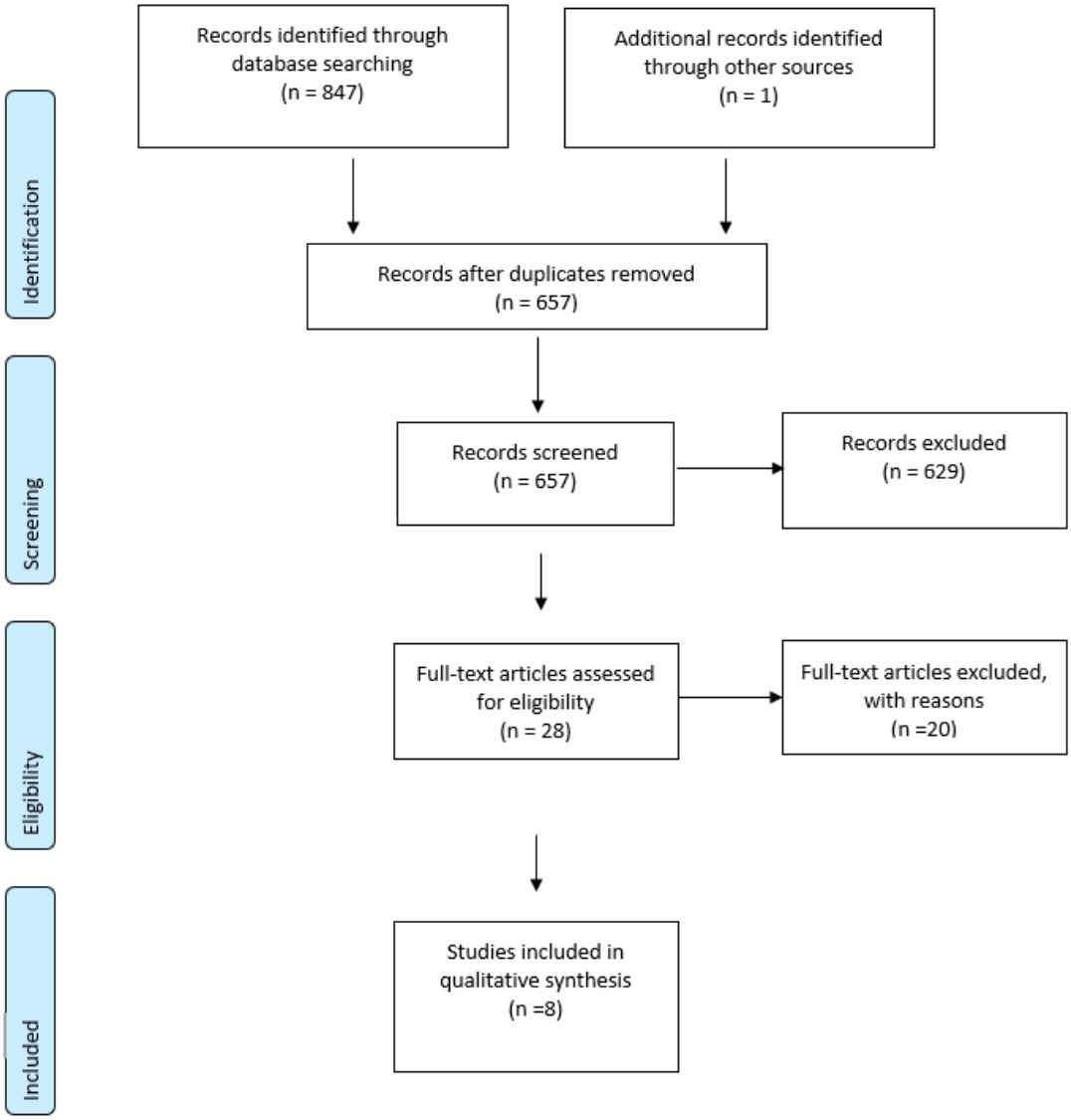
This grading does not necessarily mean a judgement that there is evidence of no relationship. With further good quality research, any exposure graded in this way might in the future be shown to increase or decrease the risk of the health outcome under investigation.

Results

Literature database search

The total number of articles identified is presented in the PRISMA [96] diagram below (Appendix Figure M-1). The database search identified 847 articles. An additional 1 article was identified through other sources. A total of 657 records remained after removing duplicates. After screening of title and abstract, 629 records were excluded and 28 remained for full-text analysis. A further 20 studies were excluded with reason after full-text analysis, leaving a total of 8 studies [30, 58-64] for data extraction.

Reasons for exclusion were; study did not meet the inclusion criteria for population (n=1), studies excluded based on outcome criterion (n=2), and study exposure variable did not meet our inclusion criteria (n=14). Additionally, three studies that were systematic reviews of reviews were excluded as they included various studies that already met our inclusion criteria (n=3). Appendix N provides a list of the studies excluded after full-text analysis and specific reasons for exclusion of each study. The quality scores for each of the included studies are presented in Appendix O. Using the AMSTAR criteria [97], five of the included studies were scored as high quality reviews [30, 58, 60, 63, 64], two were moderate quality reviews [59, 61] and one was assessed as low quality [62].



Appendix Figure M-1 Prisma flow diagram for physical activity and health outcome, all-cause mortality

Results

Strength of association

The eight systematic review and meta analyses studies included in our qualitative synthesis provided us with sufficient evidence supporting the association between PA and ACM [30, 58-64]. All eight studies included only prospective cohort studies in their analyses. Where the authors presented pooled effect sizes, we summarize the measures of disease association in our findings (Appendix table M-3). In this section, we present a narrative synthesis of our findings on the strength of association between PA and ACM.

In a systematic review and harmonized meta-analysis study, Ekelund and colleagues [30] investigated the dose-response associations between accelerometry measured PA and ACM. They found that higher levels of total PA were associated with substantially reduced risk for premature mortality Also, a non-linear dose-response pattern was evident in their findings. Hupin and colleagues [59], carried out a systematic review and meta-analysis to determine whether moderate-to-vigorous-intensity physical activity (MVPA) lower than the current PA recommendations was effective in reducing mortality. They found that when compared with participants in the inactive group, participants with a low dose of MVPA (1–499 MET-min per week), had a 22% lower mortality risk MVPA beyond this threshold improved these benefits in a linear fashion. Lollgen and colleagues [61] reported lower ACM for active individuals. A dose response relationship was evidenced by a marked reduction in mortality with light and moderate activities, and a small additional risk reduction seen for vigorous exercise intensity. A systematic review and meta-analysis by Woodcock and colleagues [64] found that being physically active reduced the risk of ACM. These authors highlight that the largest benefit was found from moving from no activity to low levels of activity. In their review, Nocon, Hiemann [62] also concluded that being physically active reduced the risk of ACM. The review study by Samitz and colleagues [63] quantified the relationships between ACM and different domains of PA. The authors found that higher levels of total and domain-specific PA were associated with reduced ACMTwo studies reported associations for walking as PA exposure [58, 60]. Of the two studies, one investigated the association between cycling and ACM [60]. These two studies reported inverse relationships between the exposures (cycling and walking) and ACM. Additionally, Kelly and colleagues [60] presented evidence of a dose-response that suggested decreasing rate of benefit at higher PA exposure. The measures of associations reported in these studies are summarised in Appendix table M-3.

For the association between PA and ACM, three sets of relative risk measures were selected for use in our model. The first set was taken from the paper by woodcock and colleagues [64]. We used the author's central estimate of moderate PA exposure levels (Appendix table M-3). We contacted James Woodcock seeking additional data used to quantify the associations between PA and ACM in their paper. In response, the author recommended a population-based prospective cohort study by Arem and colleagues [65]. Arem and colleagues used pooled data from 6 cohorts in the National Cancer Institute Cohort Consortium (baseline 1992-2003) to quantify the dose-response association between LTPA and mortality. They also sought to define the upper limit of benefit or harm associated with increased levels of PA. The authors reported that compared with no baseline LTPA, any level of activity was associated with a significantly lower risk of mortality). These values were proposed for use in our model.

Also, we selected to include measures from the study by Ekelund and colleagues [30] that examined the dose-response associations between accelerometer assessed total PA and ACM.

Appendix table M-3 Summary of the measures of disease association

Author, Date: health outcome	**Findings**
All-cause mortality	
Ekelund et al., 2019 [30]	• Total physical activity[a] was associated with lower risk of mortality with a non-linear dose-response. Hazard ratios, 1.00 (referent) in the first quarter (least active); second quarter[b], 0.48 (95% CI, 0.43 to 0.54); third quarter[c], 0.34 (95% CI, 0.26 to 0.45); fourth quarter[d] (most active), 0.27 (95% CI, 0.23 to 0.32) (n=8).

Hamer & Chida, 2008 [58]	• The pooled hazard ratio for ACM and the highest walking category[e] compared with the lowest was 0.68 (95% CI, 0.59 to 0.78) (n=10). • The pooled hazard ratio for ACM and minimal walking levels[f] compared with the referent category (using the same studies as for the main analyses), was 0.80 (0.71 to 0.91) (n=10).
Hupin et al., 2015 [59]	• Multivariate-adjusted all-cause mortality relative risks (RR) for participants of the low, medium and high moderate-to-vigorous intensity physical activity (MVPA) groups compared with participants in the inactive group were reported as below (n=9); Low-dose (1–499 MET-min per week) • RR 0.78, 95% CI 0.71 to 0.87 Medium dose [150 min of MVPA (500–999 MET-min) per week] • RR 0.72, 95% CI 0.65 to 0.80 High-dose (well above current recommendations, ≥1000 MET-min/wk) • RR 0.65, 95% CI 0.61 to 0.70
Kelly et al., 2014 [60]	• An association between 11.25 MET hours/week of walking and ACM was reported, Relative risk, 0.89 (95% CI 0.83 to 0.96) (n=14) • An association between 11.25 MET hours/week of cycling and ACM was reported, relative risk, 0.90 (95% CI 0.87 to 0.94) (n=7)
Lollgen, Bockenhoff, & Knapp, 2009 [61]	• Reported results of the multivariate-adjusted relative risk estimates for both sexes, for studies with three levels of activity were; moderate levels, 0.78 (95% CI, 0.61 – 1.00), and vigorous levels, 0.80 (95% CI, 0.66 – 0.97) (n=3). Ref group was inactive.
Nocon et al., 2008 [62]	• Relative risk of all-cause mortality in physically active versus physically inactive participants for fully adjusted models, were reported as 0.67 (95% CI 0.63, 0.72) (n=33).
Samitz, Egger, & Zwahlen, 2011 [63]	• Results on total physical activity (leisure time PA, routine activities of daily living and occupational PA) were; mortality from all causes in individuals with highest compared with lowest levels of total physical activity, adjusted RRs, 0.65 (95% CI 0.60-0.71) (n=21) • Below are overall, maximally adjusted RRs of all-cause mortality associated with an increment of 2, 4 and 7 MET-h in total physical activity per day compared with lowest level of activity (the median level for the lowest activity category was 27.3 MET-h/day; twenty-four MET-hours correspond to sitting quietly for 24 h) (n=6). 2 MET-h/day (~850 MET-min/week) 0.95 (95% CI, 0.93–0.96) 4 MET-h/day (~1800 MET-min/week) 0.90 (95% CI, 0.87–0.92) 7 MET-h/day (~3000 MET-min/week) 0.83 (95% CI, 0.79–0.87)

<table>
<tr>
<td>Woodcock, Franco, Orsini, & Roberts, 2011 [64]</td>
<td>
The association between hours of moderate physical activity and relative risk for all-cause mortality was reported as follows;
<table>
<tr><th>Hours/week</th><th>Moderate activity*</th></tr>
<tr><td>0</td><td>1.00 (Referent)</td></tr>
<tr><td>1</td><td>0.84 (95% CI, 0.81-0.88)</td></tr>
<tr><td>2.5</td><td>0.81 (95% CI, 0.76-0.85)</td></tr>
<tr><td>5</td><td>0.77 (95% CI, 0.73-0.82)</td></tr>
<tr><td>7</td><td>0.76 (95% CI, 0.71-0.81)</td></tr>
<tr><td>10</td><td>0.74 (95% CI, 0.68-0.79)</td></tr>
<tr><td>14</td><td>0.72 (95% CI, 0.66-0.78)</td></tr>
</table>
</td>
</tr>
</table>

[a] Total physical activity measured in counts per minute (cpm). [b] prespecified knot at the 25th centile of the exposure variable distributions using the medians of the quarters to define the exposure levels = 168cpm. [c] prespecified knot at the 50th centile of the exposure variable distributions using the medians of the quarters to define the exposure levels =256cpm. [d] prespecified knot at the 75th centile of the exposure variable distributions using the medians of the quarters to define the exposure levels =335cpm.

[e]highest walking exposure groups averaged more than 5.2 hours per week or more than 17.2 km per week- with considerable variation between studies.

[f]average walking time/distance in the minimal walking categories = approximately 3 hours per week (ranging from ,30 minutes per week to ,5 hours per week) or 9.8 km per week (ranging from ,5 km per week to ,15 km per week), which equated to a casual or moderate walking pace of approximately 3 km per hour.

*n activity of approximately 4.5 METs such as walking at 5.6 kilometres per hour carrying less than 11 kilograms

Abbreviations: CI, confidence interval

Appraisal of evidence and assessment of a causal relationship

In this section, we present the appraisal of evidence and assessment of causal relationship physical activity and all-cause mortality. Eight systematic review studies [30, 58-64] investigating the association between physical activity (PA) and all-cause mortality (ACM) met the inclusion criteria for our qualitative synthesis. Additionally, as mentioned earlier in our report, a population-based prospective cohort study by Arem and colleagues [65] was included in the appraisal of the strength of association and possible causal relationship between PA and ACM.

In a recent systematic review and harmonized meta-analysis, Ekelund and colleagues [30] sought to examine the association between accelerometer measured physical activity and sedentary time and all-cause mortality. The authors used individual level data from eight cohort studies, with a median follow-up of 5.8 years (range 3.0-14.5 years) and 2149 (5.9%) deaths. For our study, we focused on only findings for the association between PA and ACM.

Of the included eight studies included by Ekelund and colleagues, three were from nationally representative surveillance systems. The large sample size allowed the authors to carry out a meta-analysis of the dose-response relations between various intensities of physical activity and all-cause mortality. It was also seen to have enabled the authors to provide more precise results with narrower confidence intervals than in previous studies.

In our systematic review of reviews study, this was the only study that used accelerometry measured PA. This eliminated recall and social desirability biases that are related to self-reported measures.

Another strength seen in this study was that the authors of the primary studies included in the Ekelund review reprocessed their individual participant data according to a standardized protocol prepared by Ekelund and colleagues [30]. The authors applied a common and standardized definition of wear time, inclusion criteria, and definitions of thresholds for physical activity intensity. This reduced the heterogeneity in the cleaning and processing of accelerometer data among the studies. However, the authors note that since physical activity was only measured once, changes in PA behaviors may have affected the observed associations. Ekelund and colleagues assessed the risk of bias using the Newcastle Ottawa quality assessment scale. This is a semiquantitative scale that uses a star system to assess the quality for eight items across three domains (selection, comparability, and exposure) [30]. The quality scores for included studies were high.

Associations between physical activity variables with mortality were analyzed with three levels of adjustment. All studies adjusted for smoking. Other covariates adjusted for included wear time, age, sex, socioeconomic status, pre-existing illness, mobility limitations, and body mass index. Nonetheless, the authors caution that residual confounding may still have existed.

Compared with the referent, any level of physical activity regardless of intensity was associated with a substantially lower risk of mortality, with non-linear dose-response [30] (Appendix table M-3). To minimize bias from reverse causation, the authors excluded all deaths within the first two years in sensitivity analyses. The hazard ratios remained materially unchanged. When compared with other studies, the observed magnitude of risk reduction in this study by Ekelund and colleagues [30] is more than twice as large as previous studies that assessed PA by self-report. Their findings also extend previous studies that applied the device assessed PA by presenting findings with reduced uncertainty in the effect estimates in a much larger and more heterogeneous sample. The authors concluded that their findings provided clear scientific evidence that higher levels of total physical activity, regardless of intensity level were associated with a lower risk for premature mortality, with evidence of non-linear, dose-response patterns in middle aged and older people.

The second included study was by Hamer and Chida [58]. This was a meta-analysis of 18 prospective cohort studies with a mean follow up of 11.3 years. The authors aimed to quantify the association between walking and the risk of cardiovascular disease and all-cause mortality in healthy men and women. The inclusion of only prospective cohort studies helped eliminate recall and selection bias. Of the included studies, thirteen studies examined cardiovascular disease while 10 studies recorded all-cause mortality. For our review of review study, we only focused on the findings for all-cause mortality.

Most studies examined self-reported walking time (n=14), while other studies examined recorded walking distance (n=4), walking pace (n=6), and energy expenditure (n=2). There was considerable variation in the walking categories presented in the studies. The studies were assessed for quality based on the accuracy of the self-reported physical activity questionnaire, evaluation of outcome, and adjustment for potential confounding, with a total score ranging from 0 to 7. The total quality scores for each study ranged from 3 to 6, averaging 5. Sensitivity analyses based on study quality scores comparing high quality studies with a score of 6 with lower quality studies with a score of less than 6 did not yield different results.

The authors attempted to exclude the possibility of residual confounding by using risk estimates from multivariate models that had the most complete adjustment for potential confounders. All the included studies adjusted for age, smoking, and alcohol consumption. The authors found limited adjustment for certain confounders. For instance, only five studies adjusted for both adiposity and other physical activities. However, sensitivity analyses comparing results from these studies with the other studies gave no notable difference.

Hamer and Chida [58] found that the pooled hazard ratio of ACM in the highest walking category compared with the lowest was 0.68 (0.59 to 0.78). The effects seen were stronger for self-reported walking pace than walking volume. Also, the authors found evidence of a dose– response relationship for both walking pace and volume across the highest, intermediate, and lowest walking categories in relation to ACM. However, there was significant heterogeneity for the effect of walking on ACM. Aspects of the dose response relationship were seen to be inconsistent with other findings and the authors discuss this in their findings. Additionally, the authors caution that their analyses suggested the presence of unpublished negative findings perhaps due to lack of consistent studies investigating self-reported walking. Nevertheless, they conclude that the present findings largely support the inverse association between walking and ACM. Reduced risk of cancer was considered to partly explain the association between walking and the risk of all-cause mortality.

The systematic review and meta-analysis by Hupin and colleagues [59] aimed to determine whether a dose of moderate-to-vigorous physical activity (MVPA) below the recommended level was effective in reducing mortality. The authors included a total of nine cohort studies, with a mean follow-up of 9.8 years (±2.7) years. The methodological quality of each study was assessed using two methodology checklists. The authors considered that this had led to the inclusion of only studies with high-quality methodology, standardization of the reading of the articles, and minimized potential bias without loss of information. Studies were excluded as soon as they were not considered of high quality by both checklists.

MVPA in the selected studies was measured using questionnaires. Five of the questionnaires had been validated in adults while one had been associated with objective measures of MVPA. The PA exposure measures from each study were converted into MET-min of MVPA per week using the Ainsworth's compendium of physical activity [29]. The authors grouped the weekly MVPA into four categories: physical inactivity such as no activity beyond baseline activities of daily living (0 MET-min), low physical activity (1–499 MET-min), medium physical activity (500–999 MET-min) and high physical activity (≥1000 MET-min). The 'dose' of PA was determined by three components of physical activity: intensity (Metabolic Equivalent of Task), duration (minutes per week or per day), and frequency (days per week).

Their results showed that a low dose of MVPA (1–499 MET-min per week) led to a 22% reduction in mortality (RR=0.78, 95% CI 0.71 - 0.87) [59]. For older adults who followed the current PA recommendations, a 28% reduction in all-cause mortality was seen (RR=0.72, 95% CI 0.65 - 0.80) and a 35% reduction was seen beyond 1000 MET-min per week (RR=0.65, 95% CI 0.61 - 0.70). In their paper, the authors discussed similar findings from other meta-analyses.

Their findings suggested a curvilinear dose effect relationship. Out of the nine cohorts included, 7 cohorts reported an increase in the magnitude of the correlation with increasing doses of MVPA. Two cohorts found higher mortality risk in participants who engaged in a medium dose of MVPA than for those who practiced a low dose of MVPA. They found a steep initial slope with the greatest benefit seen for those moving from least or no MVPA to doing more. Thereafter from a medium dose to a high dose of MVPA, the relationship was linear with a smaller increase in health benefits at the highest doses of activity [59].

The authors reported having adjusted for confounding variables. To minimize the variability within single studies and between different studies, they used a random effects statistical model that had two homogeneity tests. Thy caution that the grouping of participants into four doses may have introduced a bias in the analysis due to some uncertainty in the comparison of the doses of PA. The original studies had included 3–5 dose groups. Overall, the authors concluded that MVPA reduced all-cause mortality in older adults.

Kelly and colleagues [60] conducted a systematic review and meta-analysis on reduction in all-cause mortality from walking and cycling and the shape of dose response relationship. From the 18 prospective cohort studies included, 14 studies presented a total of 18 results on walking and ACM, and 7 studies presented a total of 8 results on cycling and ACM. To assess the dose response relationship, the authors included results from 11 walking studies and 6 cycling studies.

The authors assessed for study quality using the Newcastle Ottawa Scale (NOS) for cohort studies. The NOS assessed: 1) representativeness of cohort, 2) selection of non-exposed cohort, 3) ascertainment of exposure, 4) demonstration that co-morbidities were not present at the start of study, 5) control for main variable (age), 6) controls for any additional factors, 7) how outcome was ascertained, 8) follow-up time sufficient for outcome (5 years), and 9) percentage follow-up adequate. A study could score one point in each category with a possible total of nine points. The authors found that both the walking and cycling studies were generally high quality. However, they found that the measures of walking and cycling in the included studies were often varied and crudely defined. All measures had been self-reported. Still, the authors highlighted that self-reported walking and cycling would be considered relatively valid at the group level. The authors converted reported exposure categories to MET. hours per week and derived point estimates for risk reduction at 11.25 MET-hours per week. To convert the exposures to the same metric, the authors made reasonable and transparent assumptions about the intensity of behaviors and width of category ranges. Nonetheless, they acknowledge that the assumptions and conversions may have introduced some error.

The findings by Kelly and colleagues showed that walking (18 results from 14 studies) and cycling (8 results from 7 studies) reduced the risk of all-cause mortality, adjusted for other PA. For a standardized dose of 11.25 MET-hours per week (or 675 MET. minutes per week), the reduction in risk for ACM was 11% (95% CI, 4 -17%) for walking and 10% (95% CI, 6 - 13%) for cycling [60]. The dose–response analysis showed that walking or cycling had the greatest effect on risk for ACM in the first (lowest) exposure interval with decreasing rates of beneficial effects seen as the exposure to walking or cycling increased.

The consistency of results across the included studies that investigated different exposure levels, in different walking and cycling environments supports the overall conclusion of a beneficial effect of these forms of physical activity. The authors found no evidence of publication bias. However, they did not completely rule out this possibility as it may have been possible that cohort studies finding non-significant or non-beneficial results may not have published their findings. Considerable heterogeneity was seen especially for walking. The authors explained that heterogeneity may have been due to differences in exposure categorization, sample bias, or due to residual confounding. Overall, the authors concluded that their analysis showed that walking and cycling have population-level health benefits.

In their study, Lollgen and colleagues [61] updated a meta-analysis to investigate the effect of physical activity, with different intensity categories, on all-cause mortality. They included 38 prospective cohort studies with 4 to 40 years follow up time (median of 12 years). The included studies had physical activity classifications mostly described in Kcal or METs. Though in most studies, standardized and reliable questionnaires were used to measure PA, Lollgen and colleagues point out that classification of activity and varying levels of intensity were difficult to compare and prone to different interpretation. This may have increased the variability of results.

The analysis by Lollgen and colleagues yielded a multivariate-adjusted risk lowering in males of 19 %, RR 0.81 (95 % CI, 0.75, 0.87) for moderate activity and a relative risk of 0.78 (95 % CI, 0.72, 0.84) for the most active (vigorous) group in studies with three levels of activity. Similar results were obtained for women with a relative risk of 0.76 (95 % CI, 0.66, 0.89) or a risk reduction of 24 % for moderate

activity and a relative risk of 0.69 (95 % CI, 0.54, 0.89) or a risk reduction of 31 % for the most active group [61]. A dose-response curve was seen especially from sedentary subjects to those with mild and moderate exercise. Only a minor additional reduction of risk with further increase in activity level.

An adjustment was done mostly for age, smoking status, blood pressure, cholesterol level, body weight or BMI, lung disease (in some studies) and, general health status. All the studies eligible for meta-analysis had either age-adjusted, multivariate-adjusted, or both types of relative risk estimates with corresponding confidence intervals reported. Some studies (n=8) conducted additional analyses omitting the first years (range: 1 – 12 years) after physical activity assessment to minimize potential bias from ill health in the starting population. But the results from these subgroup analyses did not have a major effect on the associations found in the total cohorts in all the studies. The authors assessed that there may have been a tendency for publication bias as positive studies may be preferred for publication. However, they note that publication bias may have been small or could be neglected because some included studies in their analysis had negative results.

The authors note that overall, the risk reduction in this analysis was comparable to those results in previous studies and reviews. They concluded that their findings presented compelling evidence of positive associations between physical activity and lower rates of mortality independent of age and sex [61]. In this study, a non-linear relationship was evident as light and moderate activities showed a marked reduction in mortality, with only a small additional risk reduction seen with vigorous exercise intensity.

The systematic review and meta-analysis by Nocon and colleagues [62] aimed to summarise the results of the largest cohort studies that examined the effects of physical activity on cardiovascular and all-cause mortality. A total of 33 prospective cohort studies with a follow up time of 4 years to over 20 years were included.

A total of nine studies used a fitness test (usually a treadmill test) to assess physical activity, and 24 studies used patient questionnaires. The authors identified that the different methods used to assess and classify PA in the included studies explained most of the heterogeneity seen in their review. From the authors' report, participants likely overestimated their levels of physical activity in self-reports, thus minimizing the true protective effect. However, they noted that even studies based on self-reported measures of activity showed marked protective effects on mortality. Additionally, the inclusion of only large studies with more than 5000 participants in this review, may have resulted in selection bias.

Though the risk reductions varied considerably, the majority of studies reported that physical activity had protective effects on all-cause mortality. This was even after adjusting for other relevant risk factors. An overall pooled risk reduction of 33% was reported [62]. The largest reductions were found in studies that used a fitness test to assess physical activity. Most studies reported results adjusted for other known risk factors such as hypertension, high cholesterol, and obesity, and age.

The systematic review and dose–response meta-analysis of cohort studies by Samitz and colleagues [63] sought to quantify relationships between all-cause mortality and different domains of physical activity. The authors included a total of 80 cohort studies with a follow-up duration of at least 2 years. A meta-analysis was done from 21 cohort studies while the dose response analysis was done from 33 studies.

To measure PA exposure, a total of 35 studies used detailed PA questionnaires, 43 used one to four questions or a brief global physical activity questionnaire and for 2 studies it was unclear what

questionnaire had been used. Many studies used ordinal categories (e.g. 'inactive', 'moderately active', 'highly active'), others used more objective criteria (e.g. MET-hours, kilocalories). Hence, substantial heterogeneity between the results from the different studies was seen. The authors opted not to convert different measures or units to one common measure (e.g. MET-hours per week).

The authors considered that the self-reported data may have been susceptible to recall bias and inaccuracies in the measurement of physical activity may have led to non-differential misclassification and attenuation of associations. Also, the assessment of physical activity only at baseline may have introduced bias.

The authors based their assessment of study quality on whether or not the study was population-based (a representative sample of the population under study), whether participants had been selected randomly, whether characteristics of study populations were clearly described (with respect to age, sex, racial or ethnic affiliation, health status, physical activity, cardiovascular risk factors, and education), whether a clinical examination had taken place before study onset and whether follow-up was near-complete. They also assessed whether analyses had been adjusted for specific potential confounding factors. Most included studies had adjusted for age, three-quarters of the studies adjusted for cigarette smoking, and half of the studies adjusted for BMI and blood pressure. Other confounders such as diabetes mellitus, lipid factors, and alcohol consumption were considered in less than half of the studies, and measures of socioeconomic status and marital status were included in less than one-third of studies. Adjustment in maximally adjusted analyses ranged from 2 to 23 variables, with a median of 7 variables. The authors considered that the length and loss of follow-up, and to what extent studies adjusted for confounding factors had both contributed to heterogeneity in their study.

In their findings, the strongest associations between physical activity and mortality were observed for total activity (RR 0.65; 95% CI 0.60–0.71), exercise and sports (RR 0.66; 95% CI 0.61–0.71) and physical activities of daily living (RR 0.64; 95% CI 0.55– 0.75) [63]. Combined RRs comparing highest with lowest categories of physical activity from the 33 studies included in the analyses were identical with the results from all 80 studies: 0.71 (95% CI 0.67–0.75) and 0.71 (95% CI 0.68–0.73), respectively. The findings suggested a curvilinear relationship between physical activity and all-cause mortality with larger benefits seen when moving from little activity to low levels of activity. Smaller additional benefits were seen when the same increment was added to higher levels of activity.

The authors discussed several biological mechanisms that would contribute to the reduction in the risk of premature death associated with physical activity. This included the favorable changes in cardiovascular risk factor profiles and improvements in endothelial function that result from PA. They presented that reductions in cancer mortality seen with increased PA may be related to reduced fat stores, increased energy expenditure, changes in sex hormone levels, improved immune function, reductions in insulin levels and insulin-like growth factors, and reduced generation of free radicals. Further, in elderly people, they noted that regular physical activity reduced the risk of falls, of osteoporotic fractures and disability, which in turn was likely to reduce mortality.

The final systematic review and meta-analysis study included in our review was by Woodcock and colleagues [64]. These authors quantified the dose–response relationship of non-vigorous physical activity and all-cause mortality. A total of 22 prospective cohort studies were included in their review.

The authors assessed the studies using the Newcastle Ottawa Scale. They assessed three areas: the selection of exposed and unexposed participants; the comparability of the groups; and the assessment of the outcome. Out of a maximum of nine stars, the authors found that the median and mean number

of stars awarded to a study was six. Two studies scored highest (eight stars) and three studies scored the lowest (four stars).

The methods used to assess and categorize PA varied across the studies. Four studies used interviews to measure PA while the rest used a self-completed questionnaire. Only one study included a repeat assessment of PA. Due to the lack of repeat measure of PA exposure, the authors considered that there may have been a high probability of unrecorded change in exposure over time, given the length of follow up (25 years in the longest study).

Using the data available in the studies and applying estimates from the compendium of activities [29], Woodwock and colleagues converted exposure measures from each study into MET-hours of activity per week. The authors noted a considerable variation of adjustment for potential confounders. All studies controlled for smoking, using different variables. Although the review authors adjusted all the studies for multiple potential confounders, they reported that potentially important confounding differences that could substantially affect the results may have remained (such as in dietary factors).

The results indicated that 2.5 h/week (equivalent to 30min daily of moderate intensity activity on 5 days a week) compared with no activity was associated with a reduction in mortality risk of 19% (95% CI, 15-24), while 7 h/week of moderate activity compared with no activity reduced the mortality risk by 24% (95% CI, 19–29) [64]. A non-linear relationship was evident. The greatest benefit was seen in the process of changing from a sedentary lifestyle to low levels of activity and smaller additional benefits from higher levels of activity. The authors concluded that being physically active reduced the risk of all-cause mortality.

The study by Arem and colleagues [65] was an additional cohort study that we included in our review of reviews study. As mentioned earlier in our report, we contacted James Woodcock seeking additional data that they may have used to quantify the associations between PA and ACM in their paper [64]. In response, the author recommended the population-based prospective cohort study by Arem and colleagues. These authors carried out a detailed pooled analysis to quantify the dose-response association between leisure time physical activity and mortality. They also sought to define the upper limit of benefit or harm associated with increased levels of physical activity. Arem and colleagues pooled data from 6 cohorts that had a median follow-up time of 14.2 years (range, 0-15.2 years). This prospective design applied was considered to have minimized recall bias. In their study, self-reported physical activity measurements were collected through various approved questionnaires. The authors indicate that this self-reported physical activity, reported at a single time point, may have limited their study. However, they note that within each of their 6 cohorts, self-reported activity has indicated construct validity. Still, the authors assessed that measurement error in self-reported LTPA was likely to have resulted in attenuation of the associations observed. Additionally, the authors used absolute, compendium-derived values to assign MET-hour-per-week intensities. They caution that this may not have accounted for inter individual variation. The dates of death in the Arem study were ascertained using the National Death Index, death certificates, or medical records [65].

The authors adjusted the final models for age, sex, educational level, smoking status, cancer history, heart disease, alcohol consumption, marital status, and BMI. Compared with no baseline LTPA, their findings showed that any level of activity was associated with a significantly lower risk of mortality [65]. A 20% lower mortality risk was seen among those performing less than the recommended minimum of 7.5 metabolic-equivalent hours per week, a 31% lower risk at 1 to 2 times the recommended minimum, and a 37% lower risk at 2 to 3 times the recommended minimum. An upper threshold mortality benefit occurred at 3 to 5 times the physical activity recommendation however,

compared with the recommended minimum, the additional benefit was modest (31% vs 39%). No evidence of harm at 10 or more times the recommended minimum was found [65].

The authors found that heterogeneity between cohorts had been statistically significant for all LTPA categories. They considered that some of the heterogeneity observed between individual study results may have been explained by differences in the questionnaires between cohorts, variation in baseline age, relative physical fitness, and length of follow-up. However, additional analyses excluding each cohort showed that estimates were not unduly influenced by a single cohort. Although the authors had attempted to adjust for confounding by history of disease or other known mortality risk factors, they caution that unaccounted risk factors may have influenced their observed results. Arem and colleagues concluded that indeed meeting the recommended guidelines by either moderate- or vigorous-intensity activities was associated with nearly the maximum longevity benefit [65].

The consistent findings from the eight review studies [30, 58-64]and one large cohort study [65] included in our qualitative synthesis provide sufficient evidence that an association between PA and ACM exists. When assessed against Bradford Hill's criteria for causality we graded our findings as convincing evidence for a causal relationship. This supports the inclusion of all-cause mortality in the *NSW Active Transport Health Model*. A summary of this appraisal of evidence for a causal relationship is summarised in Appendix table M-4.

Appendix table M-4 Assessing the evidence against causal criteria: All-cause mortality

Criteria	Description
1.Temporality	All the eight review studies [30, 58-64]and one large cohort study [65] considered in our qualitative synthesis included only prospective cohort studies with long follow up periods. The evidence met the temporality criterion.
2. Strength of association	Using an adaptation of measures by Webb and colleagues [45] to guide our classification of the strength of association, the reported measures of strength of association ranged from weak association to moderately strong. When compared with other studies, the observed magnitude of risk reduction in the study by Ekelund and colleagues [30] is more than twice as large as previous studies that assessed PA by self-report. The strength of association facilitates assessment for a possible causal relationship. Weakness of an association makes the risk of alternative explanations greater but does not preclude causality.
3. Consistency	The consistency of results across all the included studies supports the overall conclusion of a beneficial effect of PA. The authors also discuss various comparable findings of risk reductions reported in previous studies.

4. Dose-response relationship	Authors in all the eight included systematic review studies found evidence of a dose response relationship between PA and ACM.
5.Biological plausibility	Various possible biological mechanisms that would contribute to the reduction in the risk of premature death associated with physical activity were considered. This included the favorable changes in cardiovascular risk factor profiles and improvements in endothelial function that result from PA. Reductions in cancer mortality seen with increased PA may be related to reduced fat stores, increased energy expenditure, changes in sex hormone levels, improved immune function, reductions in insulin levels and insulin-like growth factors, and reduced generation of free radicals. Further, in elderly people, it was discussed that regular physical activity reduced the risk of falls, osteoporotic fractures and disability, which in turn might reduce mortality.
6. Specificity	Not supported, not readily applicable.
7.Coherence	The interpretation for the association of PA and ACM does not conflict with what is known of the natural history and biology of the all-cause mortality outcome.
Assessment of grade of evidence Convincing / Probable / Possible / Insufficient	We graded our findings as convincing evidence for a causal relationship. This supports the inclusion of ACM in the *NSW Active Transport Health Model.*

Appendix N: Studies excluded after full-text analysis

	Articles excluded after full-text review	**Reason for exclusion**
Health outcomes: All-cause mortality		
1	Barry et al., 2014 [100]	• Exposure variable- cardiorespiratory fitness
2	Biswas et al., 2015 [101]	• Exposure variable- Sedentary time
3	Chastin et al., 2019 [102]	• Excluded - measures light intensity PA
4	Chau et al., 2013 [103]	• Exposure variable- Daily sitting time
5	Cooper, Kuh, & Hardy, 2010 [104]	• Exposure variable- physical capability levels
6	Cunningham, R, Caserotti, & Tully, 2020 [105]	• A systematic review of reviews. • In relation to ACM, 1 review was included in this study. This review was already considered for inclusion in our review (Hupin et al., 2015 study)
7	Ekelund et al., 2016 [106]	• Exposure - the joint and stratified associations of sedentary behaviour and physical activity • Not possible to split results
8	Fogelholm, 2010 [107]	• Results presented as joint grouping of physical activity/fitness and obesity vs. mortality and/or morbidity.
9	Karmisholt & Gotzsche, 2005 [108]	• Whole population at that age not covered- The ACM outcome assessed is ACM in patients with coronary heart disease
10	Kraus, 2019	• An umbrella review (review of reviews) • Presents findings from some of our included individual review studies (Kelly et al.,2014, Hammer at el.,2008, Samitz et al.,2011)
11	Ku, Hamer, Liao, Hsueh, & Chen, 2020 [109]	•Excluded - measures light intensity PA
12	Lacombe, Armstrong, Wright, & Foster, 2019 [110]	• Exposure is physical inactivity in combination with additional lifestyle risk behaviours (smoking, alcohol, diet, or sedentary behaviour) and results are presented as such.
13	Liu et al., 2016 [111]	• Exposure- slow usual walking speed
14	Loef & Walach, 2012 [112]	• Exposure- combined effects of healthy lifestyle behaviours • Association for PA alone cannot be estimated from combined results
15	Patterson et al., 2018 [113]	• Exposure- Sedentary behaviour (assessed independent of PA)
16	Qiu & Meng, 2005 [114]	• Does not meet inclusion criteria • Outcome- sudden death
17	Rezende et al., 2016 [115]	• Exposure- sitting time
18	Warburton & Bredin, 2017 [116]	• Outcome - multiple health benefits • A systematic review of reviews. • In relation to ACM, only 2 reviews are included in this study. These 2 are already considered in our review (i.e Hupin et al – included & Ekelund et al. 2016)
19	Yerrakalva, Mullis, & Mant, 2015 [117]	• PA is not n exposure here but investigated as a potential confounder- explores how physical activity and

		cardiorespiratory fitness influence the association of adiposity with mortality
20	Zhao, Bu, Chen, & Chen, 2020 [118]	• Exposure- sedentary time
Health outcomes: Depression, Anxiety		
1	Ahn & Fedewa, 2011 [119]	• Study investigates relationship between PA and existing depression, anxiety, children's mental health ailments
2	Cunningham, R, Caserotti, & Tully, 2020 [105]	• A systematic review of reviews. • In relation to incident depression, 1 review was included in this study. This review was already considered for inclusion in our review (Schuch et al., 2018 study)
3	Dale, Vanderloo, Moore, & Faulkner, 2019 [120]	• They focus on reduction of depression/depressive symptoms, reduced anxiety as the health outcomes variable.
4	Rebar et al., 2015 [121]	• Health outcome is reduction of depression and anxiety and not incident depression and anxiety
5	Schuch et al., 2016 [122]	• Exposure variable- cardiorespiratory fitness
6	White et al., 2017 [123]	• The study mainly looks at mental health (term mental ill-health used too) as the key outcome variable. Existing anxiety and depression- existing considered.
7	Zimmermann, Chong, Vechiu, & Papa, 2020 [124]	• Study restricted to identification of PA as a protective factor for anxiety without describing the association.
Health outcomes: Musculoskeletal diseases - Osteoarthritis, low back pain		
1	Allen & Golightly, 2015 [125]	• The paper presents a summary of evidence and no combined measure of association.
2	Alzahrani et al., 2019 [126]	• Study looks at PA interventions for the management of existing low back pain
3	Bean, Vora, & Frontera, 2004 [127]	• Focused review study
4	Bennell, Hinman, Wrigley, Creaby, & Hodges, 2011 [128]	• not a systematic review of associations between PA and outcomes
5	Booth, Roberts, & Laye, 2012 [129]	• not a systematic review. Mentioned PA and OA association for runners not representative of whole population.
6	Bosomworth, 2009 [130]	• clinical review
7	Burton et al., 2005 [131]	• Review- A summary of the European Guidelines for Prevention in Low Back Pain
8	Curl, 2000 [132]	• Not a systematic review. Article also looks at PA and the reduction of OA symptoms
9	Dean & Soderlund, 2015 [133]	• Presents a narrative review on existing pain
10	Dugan, 2007 [134]	• Study type does not meet inclusion criteria
11	Gardiner et al., 2016 [135]	• Study type does not meet inclusion criteria
12	Jones, Schultz, & Dore, 2011 [136]	• Study type does not meet inclusion criteria
13	King, Reynolds, & De Souza, 2011[137]	• Study type does not meet inclusion criteria
14	Lefevre-Colau et al., 2016 [138]	• Critical narrative review
15	Øiestad et al., 2020 [139]	• No measures of association given and no description or definition of the PA exposure
16	Semanik, Chang, & Dunlop, 2012 [140]	• Study type does not meet inclusion criteria
17	Sitthipornvorakul, Janwantanakul, Purepong, Pensri, & Van Der Beek, 2011	• Study presents a summary of 13 individual studies included in the SR. Studies reporting for general

		population are cross sectional studies. Has 2 cohort studies reporting a population of school children
18	Stevens-Lapsley & Kohrt, 2010 [141]	• Study type does not meet inclusion criteria.
19	Urquhart et al., 2008 [142]	• Results strictly focus on the potential roles of mediating factors in the relationship between PA and OA without discussion of the association of PA and OA as a stand alone
20	Urquhart et al., 2011 [143]	• Outcome - Individual joint structures at the knee
21	Vignon et al., 2006 [144]	• The PA exposure categories - they study four categories: 1) activity of daily life; 2) structured activity or therapeutic exercises; 3) sport and recreational activity; 4) occupational activity
22	Vuori, 2001 [145]	• No combined measures of disease association provided. No combined measures of associations for the relationship between PA and LBP provided. Also, majority of the retrieved studies focused on occupational workload PA.
23	Øverås et al., 2020 [146]	• Study population for the association between PA and LBP. All eight articles on working populations investigated risk of NP and/or LBP, while the two articles on clinical/mixed populations investigated prognosis of LBP (already existing LBP, clinical population not representative of whole population for either all ages or a specific age category)

Appendix O: Quality scores for included studies

	Study	**AMSTAR score**
	Health outcome- Depression	
1	Cunningham,2020 [105]	Moderate quality review
2	Mammen,2013 [47]	Moderate quality review
3	Schuch,2018 [48]	High-quality review
	Health outcome- Anxiety	
1	McDowell,2019 [49]	Moderate quality review
2	Schuch,2019 [50]	High-quality review
	Health outcome- ACM	
1	Ekelund,2019 [30]	High-quality review
2	Hamer,2008 [58]	High-quality review
3	Hupin,2015 [59]	Moderate quality review
4	Kelly, 2014 [60]	High-quality review
5	Kraus,2019 [147]	Moderate quality review
6	Lollgen,2009 [61]	Moderate quality review
7	Nocon,2008 [62]	Low-quality review
8	Samitz,2011 [63]	High-quality review
9	Woodcock,2011 [64]	High-quality review
10	Cunningham,2020 [105]	Moderate quality review
	Health outcome- Low back pain	
1	Alzahrani, Mackey, Stamatakis, Zadro, & Shirley, 2019 [51]	High-quality review
2	Shiri & Falah-Hassani, 2017 [52]	High-quality review
	Health outcome- Osteoarthritis	
1	Richmond et al., 2013 [55]	Moderate quality review
2	Hart, Haaland, Baribeau, Mukovozov, & Sabljic, 2008 [54]	Moderate quality review

Appendix P Does active transport replace other physical activity? A rapid systematic review of the evidence

Rationale for the rapid systematic review

Active transport (AT) is increasingly recognised as a very promising means of enhancing physical activity (PA) at population level, thereby reducing the burden of NCDs [5, 10, 148] However, there are concerns that some of the physical activity gained by walking and cycling for transport (alone or in combination with public transport) may be offset by less physical activity in other domains (leisure time, worktime, activity around the home). This may affect estimates of the health benefits of investments in active transport.

Objective

We carried out a rapid systematic review to establish whether physical activity resulting from uptake of active transport is additional to existing total PA levels or if it displaces physical activity in other domains.

Methods

Data Sources and search strategy

The search strategy was informed by guidelines for systematic reviews and rapid systematic reviews [95]. The review was prepared according to the Preferred Reporting Items for Systematic reviews and Meta-Analysis Protocols (PRISMA-P) 2015 statement [96]. The review covered the three concept areas: existing physical activity levels, uptake of active transport and, change of existing physical activity levels. We searched the PubMed and Embase databases. Appendix Table P-1 shows the search strategy that we used.

Appendix Table P-1 Search strategy

1. PA levels search terms "physical activity" OR “physical exercise”
2. Active transport search terms "active transport*" OR "active travel*" OR “active commut*” OR “public transport*” OR walking OR bicycling

3. Change of existing PA levels search terms "compensat*" OR "displac*" OR "offset*"
4. Search strategy **#1 AND #2 AND #3**

Inclusion and exclusion criteria

Participants/population: our review included studies that reported results for adult populations. The search was restricted to studies carried out on humans.

We included studies that reported (1) a measure of active transportation; (2) a measure of general physical activity (e.g., total physical activity, leisure-time physical activity) and (3) outcomes on changes (or association) of physical activity resulting from uptake of active transport and existing physical activity levels at the individual level or in the populations.

Study designs: we included primary studies.

Publication status: included peer-reviewed published studies whose full text is publicly available. For studies published in multiple papers, the most detailed versions were considered.

Timeline: included studies published in the year 2009 to August 2020.

Language: included studies published in the English language.

Study Records

Data management

We imported the identified studies to EndNote software. Duplicate records were excluded. The inclusion and exclusion criteria guided our study selection.

Screening

One reviewer (MW) screened the titles and abstracts of identified studies for relevance based on the set criteria. Results of the outcome of the title and abstract screening stage were discussed and

reviewed by two other reviewers (HM and LV). The two reviewers screened a sample of the titles and abstracts of the identified studies. Once a consensus was reached, we commenced the full text review stage. One reviewer (MW) screened the full texts of selected studies. The other authors (HM and LV) reviewed and confirmed the selection process. The final list of studies was discussed and agreed upon by all authors. We documented the reasons for any excluded studies at this stage.

Data extraction

All reviewers agreed on which data to extract from included studies. One reviewer (MW) extracted data from the full texts of eligible studies. Two other reviewers (HM and LV) cross-checked the data extraction variables. The data items extracted from each study meeting the inclusion criteria are shown in Box 1.

Box 1 Data extraction fields

- 1st author's last name
- Year of publication
- Study aim
- Study design
- Study population and study area or country
- Data collection methods
- Measurement of physical activity applied in the study
- Active transport measurement applied in the study
- Data analysis applied
- Study findings
- Authors' conclusion
- Any additional comments

Data Synthesis

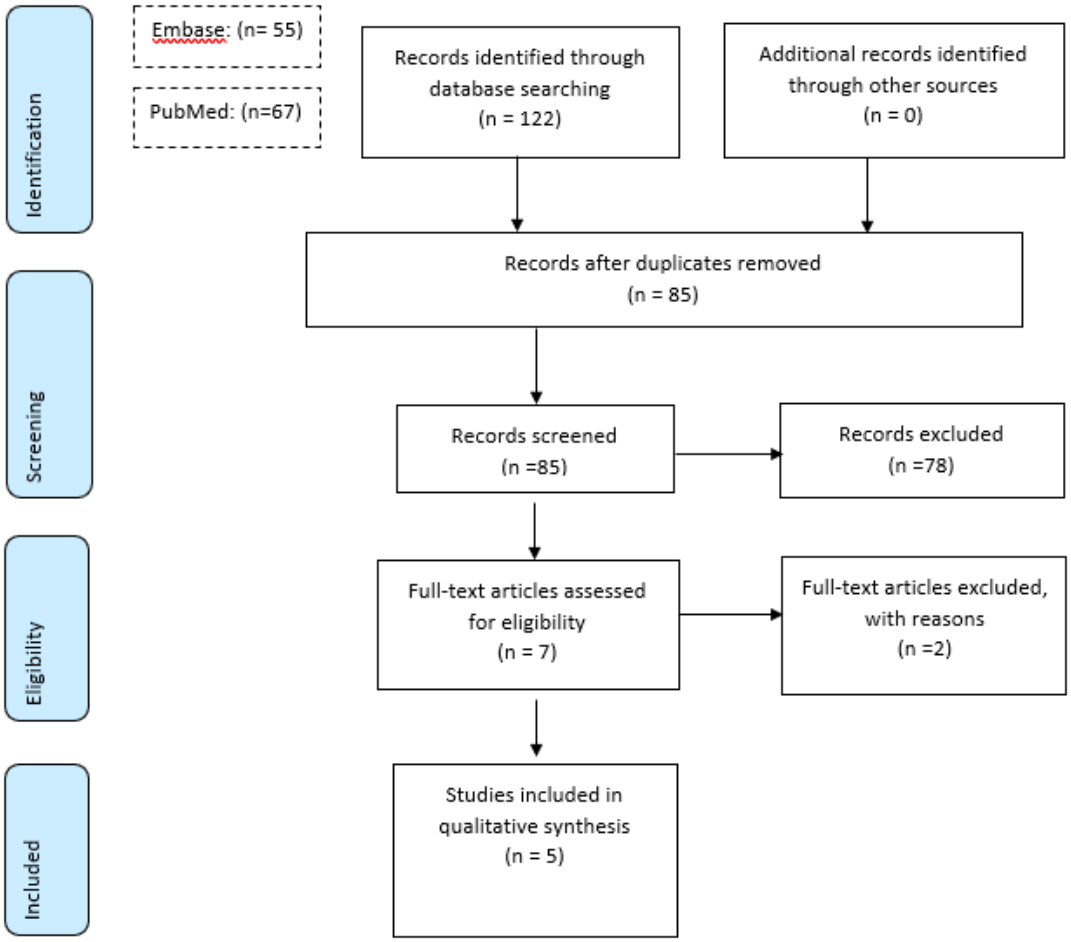
We identified studies that investigated whether additional PA resulting from uptake of active transport displaced existing PA levels. We carried out a qualitative analysis of the evidence.

Results

Selection of studies

The total number of articles identified is presented in a modified PRISMA [96] diagram below (Appendix Figure P-1). From our database search, we identified a total of 122 articles. After removal

of duplicates, a total of 85 records remained. After screening of title and abstract, 78 records were excluded and 7 remained for full-text analysis. A further two [149, 150] studies were excluded with reason after full-text review, leaving a total of 5 [151-155] studies for data extraction and qualitative synthesis.



Appendix Figure P 1 Flow diagram of our study selection process

The study by Di Blasio and colleagues [149] was excluded because it did not report on changes in PA resulting from uptake of active transport. We also excluded the study by Longo and colleagues [150] as they did not investigate the displacement of existing PA levels as a result of increased transport walking and/or recreational walking.

Study characteristics

Out of the 5 included records, two studies [151, 153] were carried out in the UK, two were conducted in Canada [154, 155] and one study [152] was done in Australia. The studies were published between the years 2014 and 2018. All studies were carried out on adult populations. Three studies [152, 153, 155] used self-reported measures of physical activity and active transport. One study used self-

reported perceived changes in various PA domains [154]. Only the study by Foley and colleagues [151] incorporated an objective measure of PA. Supplementary Appendix File P-1 gives a detailed description of the studies included in the qualitative synthesis. The studies were heterogeneous and hence we present additional details of each study as part of our qualitative synthesis of the evidence.

Data extraction and analysis

Of the five included studies, three studies [151-153] found no evidence of any displacement, meaning that active transport was found to increase physical activity and the physical activity gained by walking and cycling for transport (alone or in combination with public transport) was not offset by less physical activity in other domains (leisure time, work time, activity around the home). Results from the other two studies [154, 155] found some evidence of possible compensation of PA. Of these two, in the findings from the qualitative phase of the study by Salvo and colleagues [154], some participants described compensating their leisure time physical activity with active transportation to maintain overall physical activity levels. These participants reported to have decreased their leisure physical activity following an increase in active transportation. Findings from the cross-sectional study by Thielman and colleagues [155] suggested that higher average levels of transport walking in highly walkable neighbourhoods were partially associated with lower average levels of leisure-time physical activity in certain age and population subgroups.

The three [151-153] studies that found no evidence of compensation of PA were the strongest in our rapid review. The first study of these three was a prospective cohort study carried out by Foley and colleagues [151]. The authors examined the longitudinal associations between changes in time spent in active commuting and changes in time spent in recreational and total physical activity in adult commuters. A strength of this study was that the authors incorporated an objective measure of physical activity (heart rate and movement sensor). However, the sample size was small (n=71) and in all models, no significant associations were seen between changes in any of the active commuting exposures and time spent in objectively measured physical activity. Changes in active commuting time were associated with commensurate changes in total self-reported physical activity. In their maximally adjusted regression model, Foley and colleagues found that between the 2009 baseline and 2012, a decrease in active commuting was associated with a greater likelihood of a decrease in total physical activity (relative risk ratio 2.1, 95 % CI 1.1, 4.1). An increase in active commuting was associated with a borderline significantly greater likelihood of an increase in total physical activity (relative risk ratio 1.8, 95 % CI 1.0, 3.4) [151]. In the final model, the authors adjusted for age, sex, car ownership, access

to a bicycle, distance from home to work, baseline BMI and baseline total physical activity. Their findings indicated no compensatory change in self-reported recreational physical activity.

The second study by Gomersall et al. [152] was a randomized controlled trial that had two intervention groups and one control group. The authors investigated how previously inactive adults modified their time budgets when they undertook a new 6-week physical activity program. Their findings showed that at the end of the exercise program, participants in the intervention groups spent significantly more time in physical activity and active transport, compared to the control group. The increases in PA and AT in the intervention group, relative to controls, amounted to 21–45 min/day at the end of the intervention. The authors also found that the increases in PA and AT were largely compensated for by a significant reduction in time spent watching television and playing videogames (50–52 min/day). From these findings, Gomersall and colleagues concluded that the additional physical activity seen in the study participants was not compensated by reductions in energy expenditure in other PA domains. The third study by Goodman and colleagues [153] was a cohort study done to evaluate the effects of the Connect2 intervention on overall walking, cycling, and physical activity levels. The authors analysed the data as a natural experiment. Connect2 was an initiative established to build or improve walking and cycling routes at various sites across the United Kingdom. The authors found that living nearer the infrastructure predicted increases in activity at 2 years, relative to those living further away. At 2 years, individuals living nearer Connect2 reported significant increases in their past-week walking and cycling relative to those living further away, with an effect of 15.3 minutes per week per kilometre closer to the intervention (95% CI =6.5, 24.2). This was after adjusting for baseline demographic, socioeconomic, and health characteristics plus walking and cycling time at baseline. Proximity was also associated with a comparable increase in past-week total physical activity (adjusted effect = 12.5 minutes/week per km closer to the intervention; 95% CI = 1.9, 23.1) and not associated with any change in moderate- to vigorous-intensity activity excluding walking and cycling (adjusted effect = 0.1 minutes/week; 95% CI = –6.3, 6.5).

At 2-year follow up, the effect size was 15.3 additional minutes/week walking and cycling per km nearer; 12.5 additional minutes/week of total physical activity.

The authors hence found no evidence that the gains in walking and cycling were offset by reductions in other forms of activity. They concluded that individuals living near the infrastructure did not compensate for their increased walking and cycling by reducing their participation in other types of physical activity [153].

The evidence provided by the remaining two studies [154, 155] was weaker. One was a mixed methods sequential study conducted by Salvo and colleagues [154] among people who had moved house. The study had two aims: (1) estimate associations between changes in overall physical activity and transportation walking and cycling and changes in objectively assessed neighbourhood walkability (quantitative phase) and (2) describe perceived barriers and facilitators to physical activity following residential relocation (qualitative phase). The quantitative phase of the study relied on retrospective, self-reported perceived changes in transportation walking, transportation cycling, and overall physical activity following residential relocation were measured using a five-point scale: (1) a lot less now, (2) a little less now, (3) about the same, (4) a little more now, and (5) a lot more now. In our analysis, we considered this measurement of PA a great limitation to this study. For their qualitative phase, Salvo and colleagues carried out 14 interviews that they synthesized using a narrative informed data analysis. Walkability was measured by applying a score to change between participants' previous and current neighbourhoods. From their quantitative phase, Salvo and colleagues found that change in walkability resulting from relocation was not significantly associated with a perceived change in overall physical activity. In the qualitative phase, some participants reported decreased leisure physical activity following an increase in active transportation following residential relocation to maintain overall physical activity levels. We considered this study as the weakest in our review because the self-perceived changes to the various types of PA reported by study participants were likely to have been less accurate than the self-report or objective measures of PA.

The second study of the two mentioned above was a cross-sectional study by Thielman and colleagues [155]**.** These authors sought to estimate associations between walkability and physical activity during transportation and leisure using data from the Canadian Community Health Survey. Thielman and colleagues found that comparing highest to lowest walkable neighbourhoods, adjusted energy expenditure on transport walking was higher in the total study population, and significantly higher in all age and population centre subgroups. When comparing highest and lowest walkscore quintiles, total physical activity was higher in the subgroups; age 30–64, population centres with 100,000+ and population centres with 1000–29,999 residents. However, leisure physical activity was lower in the age 18–29 subgroup and population centres 100,000+ subgroup, but higher in the population centres 1000–29,999 subgroup. The authors concluded that their results suggested that higher average levels of transport walking in highly walkable neighbourhoods may be partially offset by lower average levels of leisure-time physical activity in certain age and population subgroups. However, they also highlighted that the higher levels of total physical activity suggested that highly walkable neighbourhoods were still associated with a net gain in energy expenditure [155].

Discussion

Of the five studies in this rapid systematic review, the three [151-153] with the strongest designs found no evidence that additional physical activity resulting from uptake of active transport displaces existing physical activity levels. Findings from the other two included studies [154, 155] were less conclusive. However, these two studies were considerably weaker in design. One relied on self-reported changes in physical activity over fairly long periods of time and is likely to be affected by recall bias [154]. The other was cross-sectional in nature and compared self-reported physical activity behaviours of people living in high and low walkable neighbourhoods [155]. The statistical analysis adjusted for the following variables: age category, sex, ethnicity, immigrant status, number of children under 12 in the household, household education, and household income quintile. This reduced the off-setting effect of leisure time physical activity by about three quarters (while having virtually no impact on the findings for transport walking), making it likely that the remaining 25% can be explained by residual confounding by measured and unmeasured variables. This suggests that the difference in leisure time physical activity may be due to differences between the characteristics of the populations living in areas with different levels of walkability that were not accounted for in the analysis, rather than due to differences in the amount of active transport they undertake.

Further evidence comes from studies that were not included in our systematic review. In a study using survey data from travellers in Baltimore and Seattle, Lachapelle and colleagues [87] found a positive relationship between transit use and leisure time physical activity, instead of the expected compensation of more active transport use by a reduction in physical activity. Using data collected from the UK-based iConnect study, Sahlqvist and colleagues [156] presented longitudinal findings on the relationship between a change in active travel and changes in recreational and total physical activity in adults. As a whole, the iConnect longitudinal study investigated the impact of newly constructed infrastructure for walking and cycling on travel, physical activity and carbon emissions. The results of the study by Sahlqvist and colleagues indicated that active travel had increased in 32% (n=529), had been maintained in 33% (n=534) and had decreased in 35% (n=565) of respondents. The changes in active travel were associated with commensurate changes in total physical activity. Compared with those whose active travel remained unchanged, total physical activity decreased by 176.9 min/week in those whose active travel had decreased and was 112.2 min/week greater among those whose active travel had increased. From these findings, Sahlqvist and colleagues concluded that an increase in active travel was associated with a commensurate increase in total physical activity and not a decrease in recreational physical activity.

Parallel evidence comes from studies into workplace physical activity. A study using Belgian time-use data from 2013 coupled with metabolic equivalent of task scores showed no association between occupational physical activity and physical activity for leisure, household work and family care, and transport, except for an association between women's occupational physical activity and physical activity in household work and family care [157]. In a cross-sectional survey carried out in Glasgow (Scotland), Tigbe and colleagues [158] used objective measures of PA to compare the 24-hour PA patterns of adults with either physically active or inactive occupations. Their results showed that there were no significant differences in physical activity between the active and inactive occupation groups, during non-work hours of workdays and during non-workdays. Having a physically active occupation was therefore not associated with lower activity, or greater inactivity, during non-work hours or non-workdays. No compensatory inactivity was seen from the participants who had a more active occupation. Another cross-sectional workplace study was carried out by Clemes and colleagues [159]. They examined sedentary behaviour and physical activity during and outside working hours in a sample of full-time office workers from England. As a secondary aim, the authors sought to investigate whether those who sat for long periods at work compensated by decreasing their sitting, or increasing their physical activity, when not at work. They found that the step counts recorded outside working hours did not vary significantly between the three workplace sitting groups (high work sitters, medium work sitters and, low work sitters). This suggested that individuals who reported greater sitting times at work did not compensate by increasing their activity levels outside working hours. In fact, participants in the highest tertile for workplace sitting reported sitting for significantly longer after work (on workdays) than low work sitters (difference = 28 min/day) and significantly greater total daily sitting times on non-workdays than low work sitters (difference = 94 min/day).

An additional finding in the literature was on the 'ActivityStat hypothesis' that was first described by Rowland [160] in 1998. The ActivityStat hypothesis suggests that when an individual increases their physical activity or energy expenditure in one domain, there is a compensatory change in another domain, in order to maintain an overall stable level of physical activity or energy expenditure. Though this hypothesis is not specific to active transport and possible changes in other PA domains resulting from uptake in AT, we included it in our discussion for additional insights on the lack of, or presence of PA displacement.

Gomersall and colleagues [161] carried out a systematic review with an aim to conceptually clarify the ActivityStat hypothesis. In this 2012 review, they also sought to examine the available experimental research with an aim to demonstrate or refute compensation in physical activity or energy expenditure. Gomersall and colleagues found that fifteen of 28 included studies provided evidence of

compensation, while 13 did not. They reported that subgroup analyses by population, type and duration of intervention, weight status and study quality also showed mixed findings. Overall, they considered the evidence that they found as inconclusive and lacking 'a cohesive approach to the question of an ActivityStat'. In 2018, Gomersall and colleagues [162] published findings of a randomised controlled trial carried out to test the ActivityStat hypothesis. Specifically, the authors investigated the effect of a six-week exercise stimulus on energy expenditure and physical activity in previously inactive adults. They measured energy expenditure and physical activity using a combination of accelerometry and detailed time use recalls using the Multimedia Activity Recall for Children and Adults. Resting metabolic rate at baseline and end-intervention was measured using indirect calorimetry. The authors randomly allocated the participants to one of the three study conditions (moderate or extensive exercise group or a control group). The trial results showed statistically significant increases in all energy expenditure and physical activity variables according to both accelerometry and time use recalls in the moderate and extensive groups, relative to controls. They found no significant change in resting metabolic rate. Considering the findings in this study, Gomersall and colleagues concluded that the results showed no evidence of an "activitystat" effect. The exercise stimuli of 150–300 min/week imposed in the study resulted in commensurate increases in overall energy expenditure and physical activity, with no sign of compensation in energy expenditure or physical activity.

Conclusion

In this study we carried out a rapid systematic review to establish whether additional physical activity resulting from uptake of active transport displaces physical activity in other domains. We have presented a qualitative analysis of the evidence identified in the five studies that we included in the review. Additionally, we have discussed further evidence from studies that were not included in our systematic review but were considered to provide additional insights in this research area. We found no convincing evidence that additional physical activity resulting from uptake of active transport displaces physical activity in other domains. We recommend further research through a comprehensive systematic review of current literature.