Experienced temperature, health and the implications for the built environment

Harry R. Kennard

A dissertation submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy
of
University College London.

Energy Institute
University College London

September 9, 2019

I, Harry R. Kennard, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

Domestic heating demand accounts for around 14% of all UK greenhouse gas emissions (BEIS, 2018c). Reduction of this demand is necessary if the UK is to meet its emissions commitments. At the same time, the consensus is that dangerous cold exposure contributes to winter mortality rates. However, determining when, where and for whom this dangerous exposure occurs is challenging. Rather than using static measures of ambient temperature, this study makes use of *experienced temperature* – a novel measure of the immediate thermal environment of an individual. The relationships between experienced temperature and sociodemographic, housing and health factors are examined using data from a longitudinal observational health study of over 100,000 participants (the UK Biobank).

Each participant wore an AX3 activity monitor for a week of everyday life between May 2013 and December 2015, which also measured temperature. The total unprocessed dataset for all participants was over 27TB. Following a considerable data processing exercise, each participant's experienced temperature and activity data were summarised in a series of metrics designed to characterise cold exposure. The resultant metrics were used in regression models against the available sociodemographic, housing and health factors to determine the relationship between cold exposure and health.

Various findings were revealed. The choice of summary metric is important to characterising the experienced temperature of a participant. The coldest times of the year are associated with lower experienced temperature for participants. Experienced temperature increases with age and decreases with activity level, health satisfaction and whether a solid-fuel open fire is used for home heating. There is clear evidence that low standard deviation of experienced temperature, named thermal variety in this study, is associated with poor health. The implications of these findings are discussed, with particular attention on who might be targeted for domestic carbon reduction schemes without risking overall population health.

Impact Statement

Within academia, the impact of this thesis is three-fold. First, from a methods perspective, it demonstrates that it is possible to measure and analyse the experienced temperature of individuals at a population level. The number of participants involved in this study far exceeds any previous studies of experienced temperature or domestic temperatures in the UK. It is hoped that this thesis demonstrates the viability of the use of wrist–worn temperature sensors in the real world. The method presented here could be of value to health practitioners and researchers who work in related fields such as fuel poverty. Second, the data outputted by the Big Data processing portion of this thesis will be available to other researchers. This opens up potential avenues for future work using the concept of experienced temperature. Finally, it contributes to the literature regarding the sociodemographic variation of temperatures in the UK. It is expected that further publications regarding the relationship of experienced temperature to health will follow in the months after the completion of this study.

Outside academia this study has the potential to contribute practical understanding which could aid the development of commercial heating or thermal comfort control systems. It is conceivable that a wrist worn device could be developed that would automatically regulate heating in order to maximise the efficiency of such systems. As the economy is decarbonised, improvements in the energy efficiency of home heating devices is essential. This is especially true if the UK adopts heating systems based on electric heat pumps, but also applies if bio-gas or hydrogen based systems are preferred. Beyond this, there is an opportunity to critically engage with policy makers regarding the immediate health impacts of adequate domestic heating for vulnerable populations, and in the longer term, the wider scale impacts of carbon based heating systems for the population as a whole.

Acknowledgements

This research has been conducted using the UK Biobank Resource under Application Number 26284. This research was made possible with support from the EPSRC Centre for Doctoral Training in Energy Demand (LoLo), grant numbers EP/L01517X/1 and EP/H009612/1, funding from UK Research and Innovation through the Centre for Research into Energy Demand Solutions, grant reference number EP/R 035288/1 and from the RCUK Centre for Energy Epidemiology EP/K011839.

The primary acknowledgements go to my supervisors Gesche Huebner and David Shipworth. Their insights and guidance were invaluable. Raquel Alegre and Stuart Grieve were essential to gaining the understanding to undertake the massive data processing task that this study required. Jane Galbraith and Rex Galbraith helped me out with understanding statistics early on. Cliff Elwell, who was an excellent MRes supervisor, made helpful suggestions in the early stages of the PhD. Jez Wingfield gave invaluable technical assistance with some of the monitoring equipment used in this study. Jenny Crawley's experience and advice was very important for getting to the end of the PhD. Comments during development phase from Mike Tipton, Wouter Marken van Lichtenbelt and Hannah Pallubinsky were valuable. Ed Sharpe suggested looking into using NASA data for the external temperature variable used in this study. The other students on the LoLo doctoral training program made the PhD much less painful than it might have been. Thanks too to all the members of the PACE research group, for their feedback and comments throughout.

All my family, in Wales, England, Scotland and Texas, deserve special thanks for supporting me throughout the thesis. Most of all I am grateful to Anna, for her patience, clarity and unshakeable confidence that I would be able to finish this thesis. Without her none of this work would have happened.

Contents

List of Figures

List of Tables

Glossary

C_{EWD} Conditions associated with excess winter deaths. 20, 21, 25–27, 94, 99, 102–104, 159–173, 175, 182, 204, 205

t_{10}^{f} The first decile of the experienced temperature filtered to only include times where the participant's activity level was below 10mg. 21, 130, 131, 197

t_{10}^{m} The first decile of the experienced temperature, filtered to only include times where the activity was below the participant's median level.. 18–21, 26, 88, 89, 91, 93, 94, 101, 117, 125–127, 131–138, 140, 145, 146, 158–160, 167–171, 177, 181, 183, 189, 197, 202

t_{μ}^{m} The mean experienced temperature, filtered to only include times where the activity was below the participant's median level.. 20, 140–144, 146, 157, 158, 170

t_{iqr}^{m} The interquartile range of experienced temperature, filtered to only include times where the activity was below the participant's median level.. 170

t_{max}^{m} The max of the experienced temperature filtered to only include times below the participant's median activity.. 158

t_{min}^{m} The minimum of the experienced temperature filtered to only include times below the participant's median activity.. 18, 19, 26, 89, 117, 134–138, 149, 150, 158, 170, 175, 178, 181, 189, 202

t_{range}^{m} The range of experienced temperature, filtered to only include times where the activity was below the participant's median level.. 158, 170

t_{sd}^{m} The thermal variety, defined as the standard deviation of the experienced temperature, filtered to only include times where the activity was below the participant's median level.. 20, 21, 24, 26, 27, 145, 146, 148–153, 158, 170–172, 175, 204, 205

t_{10}^{q} The first decile of the experienced temperature filtered to only include times just below a *moderate* activity level of 100mg. 130, 131

t_{10} The first decile of the experienced temperature.. 19, 21, 26, 117, 125–127, 131–138, 140, 146, 148, 156, 158, 170, 171, 175, 178, 181, 182, 189, 197, 202

t_{μ} The mean experienced temperature.. 20, 140–144, 146, 156, 158, 170

t_{iqr} The interquartile range of experienced temperature.. 153, 154, 170

t_{max} The maximum of the experienced temperature.. 154, 158

t_{min} The minimum of the experienced temperature.. 19, 26, 91, 117, 132–138, 149, 150, 153, 158, 170, 175, 178, 181, 182, 189, 202

t_{range} The range of experienced temperature.. 153, 154, 158, 170, 175

t_{sd} The thermal variety, defined as the standard deviation of the experienced temperature.. 20, 21, 24, 26, 27, 91, 139, 145, 146, 148–154, 156, 158, 170–172, 175, 178, 180, 182, 189, 204, 205

AX3 The open-source wrist worn activity and temperature monitor used in this study. 17, 18, 23, 38, 39, 41, 42, 63, 71–76, 79–81, 84, 86, 90–95, 104, 107–114, 119, 128–130, 136, 139, 145, 154, 167, 169, 179, 189–192

Axivity The manufacturers of the open-source AX3 activity and temperature monitor. 83

BASH The Borne-again Shell. A programming language. 84

cluster The large array of computers that allow parallel processing. The systems used in this study were called Legion and Myriad. 83–85, 192

CSV Comma separated value file. 83–85

CWA The file format in which the AX3 data is encoded during measurement. 73, 83, 84, 90, 107

df degrees of freedom. 24, 127, 128

DTR Diurnal Temperature Range. 44, 48, 145

EWD excess winter death. 36, 53–55, 159

GLM Generalised Linear Model. 103

HOBO The U12 Temperature Relative Humidity 2 External Channel Logger. Manufactured by Onset. Range: -20°C to 70°C. Accuracy: ±0.35 °C at 0°C to 5°C. Resolution: 0.03°at 25°C. 18, 107–109, 114

iButton A small temperature sensor typically used to measure skin temperature. 18, 62, 79, 80, 109–112

ICD International Classification of Diseases. 36, 76, 99, 102, 105, 129

IET individually experienced temperature. 62

Java A programming language. 83, 108

JSON JavaScript Object Notation. 84

LOESS Locally Estimated Scatterplot Smoothing. 20, 21, 119, 146, 148, 168, 172

LR Likelihood Ratio. 100, 119, 160

MATLAB An engineering programming language and computing environment. 107, 108

MLM Multilevel Modelling. 99, 100

MPVA Moderate to Vigorous Physical Activity. 130

MTA Material Transfer Agreement. 82

Mtoe Million tonne of oil equivalent. The energy contained in a million tonne oil. Equal to 11.63 MWh. 17, 34, 35

NASA MERRA-2 The dataset of surface temperature used in for the external temperature variable in this study.. 93, 128, 129

OLS Ordinary Least Squares. 126, 127, 131

OSGB Ordnance Survey National Grid. 93

PAP Pre–analysis plan. 23, 78, 79, 81, 87, 94, 97, 101, 126, 155, 178, 192

Python A programming language. 83, 85, 93, 108

R A programming language. 83, 93, 104, 160, 167

RDS Research Data Services. 84, 85

RR Risk ratio. 103, 104, 170, 171

SCP Secure copy. 84

SET Standard Effective Temperature. 62

SQL Structured Query Language. 83

UK Biobank The biobank used in this study. 19, 23, 38, 39, 70–72, 78, 79, 81–84, 87, 93, 95, 100, 104, 105, 108, 112, 115, 118, 135, 191, 193, 194

VIF Variance Inflation Factor. 127, 167, 171

VPC Variance Partition Coefficient. 101, 119, 160

WT wrist temperature. 64–68

Chapter 1

Introduction and background

Every act of energy conservation like this is more than just common sense, I tell you, it is an act of patriotism

Jimmy Carter – A Crisis of Confidence (15 July, 1979)

During the 20th century, global life expectancy rose from 31 to over 65 years (Prentice, 2008). Over the same period, per capita total energy use almost trebled (Brito and Sousa, 2015). These increases were primarily driven by economic development, and while neither directly causally linked nor equitably distributed, their relation reflects the complex ways energy use and health are interlinked. Fossil fuel energy sources provide the vastly dominant portion of this energy – 82% of global primary energy supply was fossil fuel in origin in 2016 (IEA, 2018). Their continued use is the primary driver of anthropogenic climate change (Hansen et al., 2011).

If catastrophic global warming is to be curtailed, the first half of the 21st century will be characterised by the drastic decarbonisation of the global economy. This will likely require a decoupling of many of the interactions between health and carbon intensive energy use. For example, the use of carbon energy sources has direct impacts through air pollution, from both supply (Markandya and Wilkinson, 2007) and demand perspectives (Lam et al., 2012). Global warming and its consequences are increasingly framed as a public health emergency (Levy and Patz, 2015). This thesis is situated within such a framing.

The focus of the thesis is health within the built environment. Around 40% of global CO_2 emissions and 36% of primary energy use is attributable to buildings (IEA, 2018). In 2007, The Lancet published a series of papers examining the links between energy and health. As part of this series, Wilkinson et al. (2007) laid out a number of the interconnections between health, energy and buildings. This study is motivated by questions surrounding one such link, that of temperature and health. In the broadest sense, it aims to contribute to understanding of how the temperatures we are exposed to in daily life relate to our health. In temperate climates such as the UK, home heating practices are clearly important in this regard, given that we spend around 90% of our time indoors (Brown, 1983). The following dilemma is central to this question. Space heating is a major use of energy in such climates. This energy use is essential for the maintenance of health and the avoidance of the worst effects of colder winter temperatures. At the same time, it is the use of this highly carbon intensive heating that carries its own risks in terms of global warming.

The first chapter sets out the context and motivation for the study. It describes the structure of the thesis and provides an overview of its scope and potential impact. In the next section, a more detailed examination of the present energy demand context is given, focusing primarily on the United Kingdom, where the study's data were collected.

1.1 The energy demand context

Starting with the global picture, and focusing in on the UK and the UK building stock in particular, this section lays out the energy demand context of the study. According to IEA estimates, global primary energy demand was 14,050 Mtoe (163,402 TWh) in 2017; an increase of 2.1% on the previous year and equivalent to 2.4 kW continuous power per capita (IEA, 2018).

Anthropogenic fossil fuel CO_2 emissions grew 1.2% in 2017 to 37.1 Gt CO_2 after two years of stability (Muntean et al., 2018). In more developed countries, the hitherto existing correlation between economic growth and energy demand has begun to be decoupled. Recent economic shifts in more developed countries away from manufacturing have allowed these nations to in effect export carbon emissions overseas, mainly to China (Davis and Caldeira, 2010).

In 2017, approximately 0.8% of the world's population lived in the UK and used 1.0% of the total primary energy demand (IEA, 2018). The 149.1 Mtoe (1,734 TWh) of UK demand was dominated by natural gas (30%), petroleum (45%) and electricity (14%), following a 96% reduction in coal, coke and breeze usage since 1970 (BEIS, 2018a). Domestic usage accounted for 28% of the UK total (40.1 Mtoe, 466 TWh), with space heating accounting for 63% of total domestic energy use (25.4 Mtoe, 295 TWh) and 75% of this was provided by natural gas (18.9 Mtoe, 220 TWh) (BEIS, 2019). The changes that have occurred in the UK for each sector of the economy are shown in figure 1.1.

UK total greenhouse gas emissions are falling – approximately 2% on average per year since 1990 – and were 456 $MtCO_2e$ in 2017 (BEIS, 2018c). CO_2 emissions, which comprise the vast majority of greenhouse gas emissions, were 367 $MtCO_2$ in 2017, which equates to just under 6 tCO_2 per capita. Residential CO_2 emissions have dropped 18% since 1990 levels and stood at 64.1 $MtCO_2$ in 2017 (17.5% of total CO_2). (BEIS, 2018c).

The UK is one of the few countries to have legally binding greenhouse gas reduction legislation. The Climate Change Act of 2008 was amended in July 2019 to introduce a target

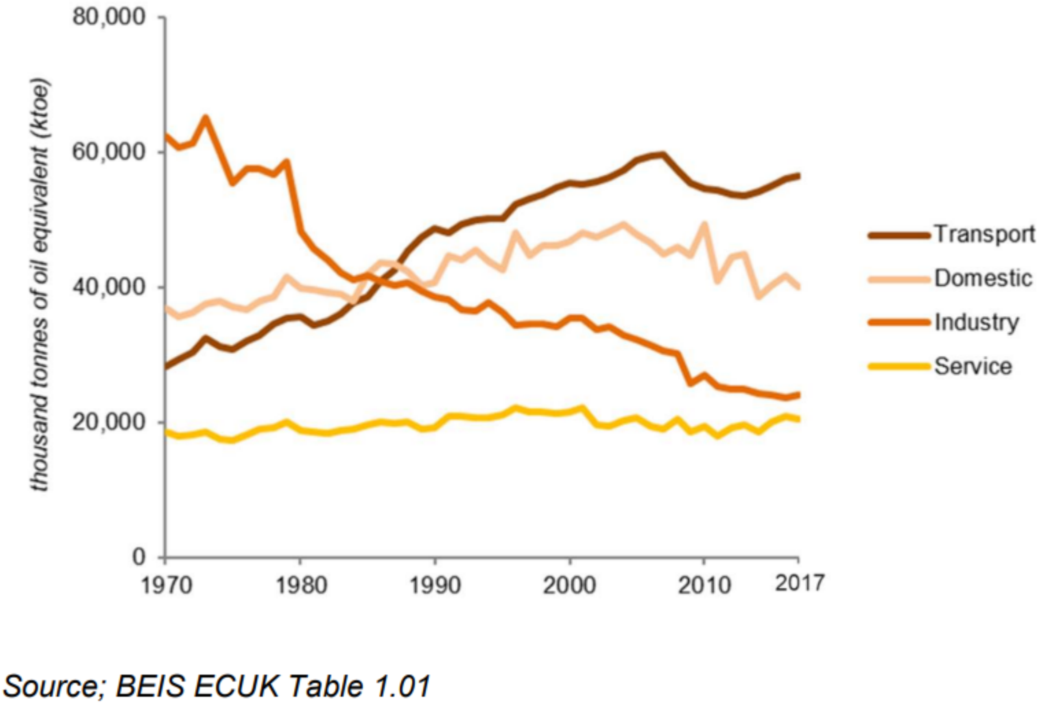


Figure 1.1: UK Annual energy consumption by sector. The UK population was 55.63 million people in 1970 and 66.04 million in 2017 (an increase of 18.7%). Total energy consumption was 156.8 Mtoe (1,824 TWh) in 1970 and 149.1 Mtoe (1,734 TWh) in 2017, a fall of 4.9%. Data source: BEIS (2018a)

of at least net-zero greenhouse gas emissions by 2050 (UK Government, 2008), which now makes it the most stringent target of any major industrialised economy.

1.2 Housing

Since this thesis focuses on temperature and health within the domestic building stock of Great Britain, it is instructive to briefly survey the character of the stock in general. There are around 28 million dwellings in Great Britain. 63% are owner occupied, 19% privately rented, 10% housing association rented and 7% rented from local authorities. 85% of British homes are in England. Nearly 56% of these were built before 1965 when building standards were introduced (MHCLG, 2019). At the time of the survey in 2017, almost a fifth did not meet the government defined Decent Homes standard (MHCLG, 2006). Over a third of homes built before 1919 and only 2% of homes built after 1990 were found to be 'non-decent'. As of 2017, 80% of buildings are classed as residential (BEIS, 2018b). 17% of people live in rural areas, and 83% in urban areas. 1.1% of people live in areas described as sparsely populated (DEFRA, 2019).

Using data derived from the English Housing Survey and other national sources, Palmer and Cooper (2013) reported on the state of domestic energy use in the UK in 2013. Overall, the housing stock changes very slowly - the number of households is increasing at a rate of less than 0.9% a year, and average household size is falling. The distribution of homes is shifting towards the South, South West and Midlands, relative to the North of England. A third of the stock is comprised of flats and detached homes, and they are increasingly common. The stock is increasingly privately owned; local authorities own four million fewer homes than in 1970. Home heating energy use increased by two-fifths between 1970 and 2013, although it fell between 2004 and 2009. The average SAP rating, which is a measure of domestic energy efficiency, is improving slowly. By 2011, 90% of homes had central heating. According to the report, modelling suggests that the average temperature in UK homes rose from 13.7°C in 1970 to 17.7°C in 2011 - although it should be noted that these averages do not necessarily reflect the temperatures experienced by inhabitants. These topics will be returned to in greater depth in chapter 2.2.1.

1.3 Cold exposure and health

As mentioned above, one of the primary reasons we heat our homes is to avoid the impacts of cold during winter. In the UK, the majority of homes use natural gas for heating. With an oceanic climate, the mean UK temperature typically ranges from around 5°C in winter to 15°in summer. Consequently, days where the temperature falls below freezing are common in winter (Kendon et al., 2018).

The primary epidemiological evidence that cold exposure is linked to health problems is the peak in mortality in the coldest months. This is demonstrated in figure 1.2, which shows how monthly numbers of deaths in England and Wales vary with age. In order to remove the effect of the background mortality rate, the number of excess winter deaths (EWDs) is typically used. This is calculated using the following relation,

$$\text{EWD} = \text{winter deaths} - \text{average non winter deaths}. \tag{1.1}$$

Winter in the northern hemisphere is normally taken to be December to March. There has been some debate as to whether measuring EWDs is the best way to characterise the peak in winter mortality (Hajat and Gasparrini, 2016). This, and related issues, are considered in greater depth in section 2.1.3.

Winter 2017/18 recorded the highest peak in EWDs in England and Wales for over 40 years. The primary causes of death in UK EWDs are circulatory diseases (ICD-10 codes I00 to I99), respiratory diseases (ICD-10 codes J00 to J99) Alzheimer's disease and dementia (ICD-10 codes F01, F03 and G30) (ONS, 2018). The International Classification of Diseases (ICD) code refers to the unique identifier of diseases produced by the World Health Organisation (the version used in this thesis is ICD-10). These conditions are exacerbated in

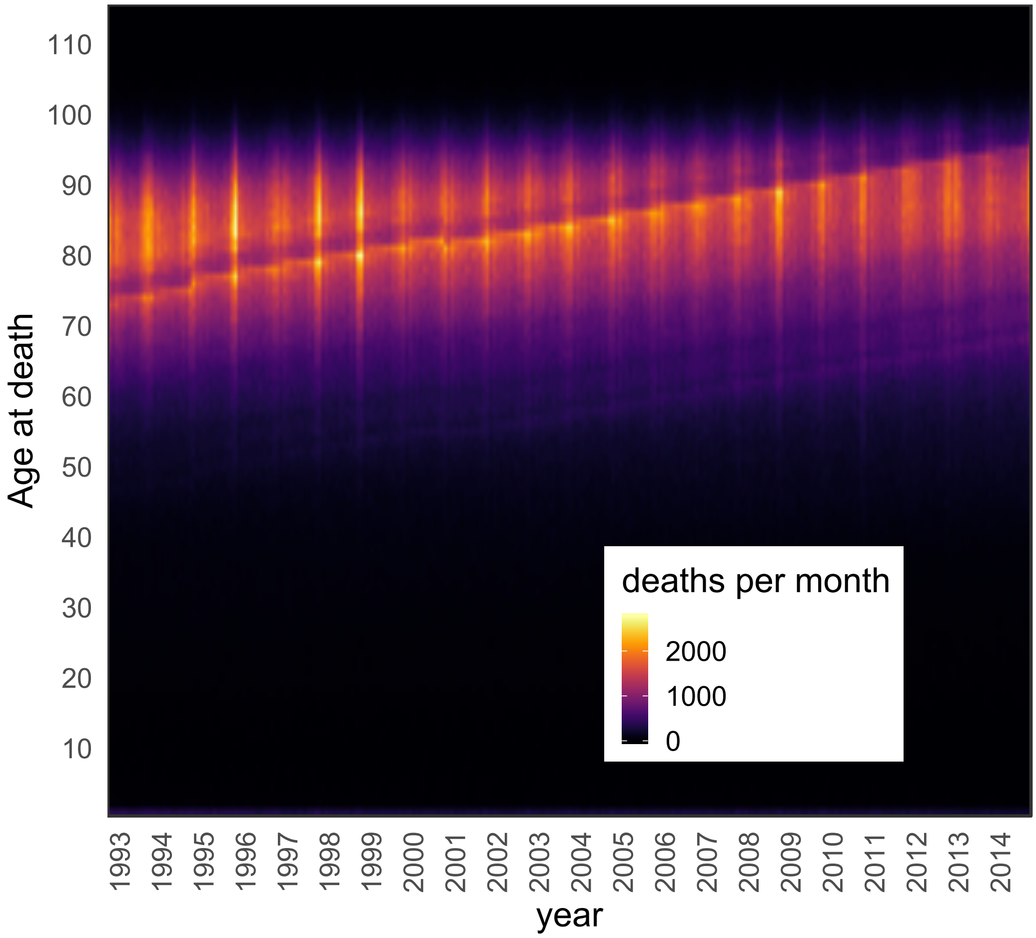


Figure 1.2: Monthly deaths by age in England and Wales at age of death. The bright peaks correspond to winter time. The diagonal features correspond to the decreased birth rate in the 1st and 2nd World Wars. Data source: ONS (2016).

the winter and are consequently correlated with lower daily average external temperatures. However, there is ambiguity as to the relative importance of long duration moderate cold exposure versus short-term more intense cold stress in the development of these illnesses. This distinction can be characterised by longer periods in a poorly-insulated home against short waits at a very cold bus stop. Keatinge and Donaldson (1997) suggest "cold stress to people waiting at a bus stop in a cold wind can exceed anything experienced indoors.". Population level studies show a non-linear relationship between external temperature and mortality (Gasparrini et al., 2010), but it is unclear what the relevant exposure function is at the level of an individual. Moreover, cold exposure studies which use ambient temperature recorded at local weather stations suffer from assuming homogeneity in temperature within a single study location, and sensors located in the home fail to account for actual exposure. Problems such as these motivate the concept of experienced temperature, where sensors record the immediate thermal environment of the participant (Kuras et al., 2015). It is the development of this concept that is one of the central concerns of this thesis.

1.4 Problem statement

The above introduction allows the research problem of this thesis to be laid out. Primarily, this thesis is concerned with defining and measuring the experienced temperature of an individual. It attempts to understand how this measured experienced temperature relates to cold exposure. In contrast to subjective determinations of feelings of cold or discomfort, it aims to understand how possible it is to define an objective experienced temperature and understand how well this characterises the immediate thermal environment of people. It then asks how the experienced temperature relates to sociodemographic, building and health factors more broadly, with a view to understanding and reducing heating energy demand.

1.5 Data source: UK Biobank

The primary data source for this thesis is the UK Biobank (Sudlow et al., 2015). It is a large-scale, longitudinal cohort health study which recruited 500,000 people in 2006-2010 aged between 40 and 69 years, principally from major urban centres across Britain. It is currently funded until 2022, but it is expected that it will follow its participants for many years to come (Sudlow et al., 2015). Each participant provided detailed baseline health and demographic information. The resource is also linked to participant health records, which allows the development of diseases and eventual causes of death to be followed. A sub-sample of over 100,000 participants wore an Axivity AX3 activity monitor continuously for a single week within the study period which ran between May 2013 and December 2015 (Doherty et al., 2017). It is this device that provides the data which are analysed in this thesis. The AX3 also measures temperature, which is used as the basis of the experienced temperature variable in this study. A substantial portion of this study involved the processing and downsampling of the AX3 data, which was over 27 TB in total. This is described fully in section 4.2.1.

1.6 Scope

As with all interdisciplinary studies, it is important to restrict the scope of the investigation, in order to prevent it expanding beyond the limits of a standard PhD. With this in mind, this study focuses on UK domestic energy demand and the health outcomes associated with cold exposure specifically. Furthermore, questions of thermal comfort and thermal sensation will not be addressed directly. Fuel poverty – the condition in which households lack the material means to afford warmth – will also not be considered directly, but aspects of both fields are relevant to the study as a whole, as will be shown in the literature review in the following chapter. While the focus of the study is Great Britain, many of its conclusions are relevant to other countries with similar climates and built environments, such as Ireland and New Zealand. While an array of Big Data computing methods have been used during the course of this thesis, they are not a primary focus of the discussion.

1.7 Relevance

Energy demand reduction as part of the decarbonisation of the economy is one of the most pressing issues that faces us, and is likely to remain so for at least the next 30 years. Achieving the net–zero carbon emissions target of 2050 will require deep restructuring of almost every aspect of daily life. Understanding how to do this as quickly as possible, whilst also minimising harm to health, is a massive sociotechnical challenge. It is hoped that this thesis contributes in a meaningful way to the understanding required to meet this challenge. It is expected that the issues discussed throughout are of immediate relevance, not just to the energy demand community, but also to physiologists and public health practitioners. Those working in the field of fuel poverty will likely find some of its methods and conclusions helpful. In principle, it could aid the development of new forms of heating control systems which focus on providing warmth for the user at the individual level, as opposed to the whole-home approach.

A key output of this study is the database of experienced temperatures that result from the processing of the AX3. As part of the data access agreement, these data will be made available to other UK Biobank researchers, allowing them to conduct their own analysis.

1.8 Thesis structure and conventions

The following chapter will comprise a comprehensive literature review of the fields briefly surveyed in this introduction. In Chapter 3, the central conceptual model underlying experienced temperature is described, as well as how the UK Biobank is structured and how the AX3 wristband used in this study. It also describes the key research questions and the hypothesised results. In chapter 4 the main method of the thesis is specified. The following chapters, (5, 6, 7 and 8) contain the results to each of the research questions in chapter 3, as well as an analysis of subsequent questions which arose during the analysis. The results as a whole are critically examined in chapter 9, before the conclusion to the thesis in chapter 10.

The thesis adopts a number of conventions when reporting results. Confidence intervals for estimates are given at the 95% level unless otherwise stated and are denoted with square brackets. Generally, the term *outcome variable* is used in a manner synonymous to dependent variable in other studies. Equivalently, *explanatory variable* is used in place of independent variable. A full list of the software used in the study is given in appendix F.

Chapter 2

Literature Review

Winter night in Harlem
Radiator won't get hot
Well the mean old landlord, he don't care
If I freeze to death or not

Bill Withers – Harlem (1971)

In the broadest sense, this thesis is motivated by questions surrounding the impacts of cold exposure on health. It takes a realist philosophical approach, and holds that it is possible to extract coherent information about these concepts from available sources to construct falsifiable claims about this information. While there are multiple critiques of the concepts of 'health' from a social scientific perspective (Nordenfelt, 2006; Kingma, 2007), and to a much lesser extent temperature and other physical variables, it is not the aim of this thesis to engage with these debates.

Figure 2.1: The three different scales which this literature review focuses on are represented by the three images above. The first represents the population scale impacts of cold exposure as demonstrated by epidemiological studies. The second image represents questions around human physiology and temperatures at the level of housing are considered. Since the study uses the wrist worn AX3, the third image represents evidence for variations in wrist temperature.

	At home (%)	In bedroom (%)	In living areas (%)
Full-time worker	49	30	19
Part-time worker	65	33	32
People who stay at home	69	33	36

Table 2.1: Estimates of the percentage of time in the year spent at home for different subsets of the UK adult population (reproduced from Daraktchieva (2018))

This chapter gives an in-depth analysis and critical review of the literature related to this thesis. The literature sits at the intersection of several fields, which will be addressed in the sections to follow. They are illustrated schematically in figure 2.1. In order to aid comprehensibility, the topics addressed are split across the three scales illustrated in the figure. It is important to note that these scales interact with each other, and are not separated - the widest scale population effects influence individuals and vice versa. From the widest perspective, the epidemiological evidence suggests strong seasonal variation in mortality. Since this phenomena occurs at a large scale, this is represented by a cityscape in figure 2.1. Evidence at this level is usually collected over long periods of time, in some cases several decades. Studies at this level usually examine broad population features such as mortality rates or disease prevalence. Consideration of this scale is essential for providing the backdrop against which more fine level effects of cold exposure occur.

Below this is the household and individual level, illustrated by the person in the central portion of figure 2.1. At the household level, low indoor temperatures have been linked to negative health outcomes for occupants, for both mental and physical health. Substantial variation in building stock characteristics accounts for some of the observed variation in domestic temperatures. It is important to consider household temperatures as people spend a large proportion of their time at home. Estimates of this are given in table 2.1. Across the whole UK population, adults are at home approximately 52% of the time. Including work and other activities, the total percentage of time spent indoors is around 90% (Brown, 1983). Diversity is also evident in the thermophysiological response to different temperatures, ranging from the observation that cold exposure increases blood pressure in older people (Collins et al., 1985) to evidence that mild cold exposure may positively benefit metabolic health (van Lichtenbelt et al., 2017).

The final level is illustrated by the third image in figure 2.1 which shows an AX3 wristband being worn. Since this thesis uses the wrist worn AX3 temperature and activity monitor, it is essential to understand both previous studies of wearable temperature monitors and the details of wrist temperature variation. Again, there are links between this level and the levels above – wrist temperature is influenced by thermophysiological factors as well as the patterns of daily life that vary throughout the population.

The review as a whole makes use of a number of different methodological approaches to literature reviews. Firstly, there is the standard approach using search engines such as Scopus and Google Scholar to locate papers and books of primary interest to the research scope. This approach is used for the majority of the papers discussed in this review. It is used when overviews of the literature are required, as it is less time intensive than the systematic review (described below). Second, a novel method which traces the references of a paper to understand the original data source of a particular claim. This method is used to assess a chapter of a widely used report in the field of the impact of cold homes on health written by Marmot et al. (2011). This method gives insight into how the bodies of evidence that are used in this field are constructed. Third, the method of a systematic review is used. Systematic reviews typically aim to encompass all existing literature on a topic. They are exhaustive and time consuming but are very useful in instances where it is important to understand the full extent of the literature covering a particular topic. This method is used in this thesis for understanding the evidence for the variation of wrist temperatures in the population.

Prior to exploring these three levels of the literature, it is it is helpful to outline some of the key temperatures which are relevant to the interactions between humans, health and the built environment. This helps to contextualise the temperatures found in the literature. Under typical conditions, humans maintain a core body temperature of around 37°C. The classical thermal neutral zone is defined as the range of ambient temperatures for which there are no regulatory changes in metabolic heat production or evaporative heat losses. For sedentary nude individuals it is between 26.8°C and 28.9°C in the steady state (Kingma et al., 2014). Clothing, activity levels and thermophysiological responses such as shivering and sweating expand the range of ambient temperatures for which humans can maintain their core body temperature. Ultimately, the maintenance of core body temperature is a key determinant of much of the material make-up of the built environment, the kind of environments we can live in and the activities which are possible in everyday life.

Under conditions far beyond the thermal neutral zone, the metabolic processes necessary to life are no longer able to be sustained, and death from hypothermia or hyperthermia results. However, in the United Kingdom, such outcomes are very rare (ONS, 2017). For reasons of scope outlined in the thesis introduction, this review is limited to the deleterious effects of cold exposure – but relevant differences to heat exposure are discussed where appropriate. Therefore, this review is mostly concerned with the evidence around what happens in between thermal neutrality and extreme cold exposure.

Jevons et al. (2016) reviewed the available evidence on indoor temperature thresholds for health. Following a systematic search of UK and similar climate evidence they concluded that there is limited evidence available on minimum temperature thresholds for homes. Despite this, they conclude that at least 18°C suffices for the population as a whole, and they suggest that this guideline should be accompanied with "nuancing of messages for those more vulnerable to the effects of cold" (Jevons et al., 2016). The papers that the authors reviewed are summarised in a separate publication (Wookey et al., 2014). Of the 20 papers included in the final review, six used blood pressure as a proxy measure of health status (see below, section 2.3). Others used body mass index and other haematological measures. The review summary points out that these are 'reasonable approaches, but are only proxies for harder health outcomes such as cardiovascular disease or health service use'. None of the papers included in the review considered mental health outcomes, which is a limitation if health is considered in broader terms.

Finally, the general temperatures of the UK climate are useful to outline. Kendon et al. (2019) provide a comprehensive overview of the UK climate in 2018. The UK climate is warming. In 2018, the mean UK temperature was 9.5°C, which was 0.6°C above the long-term 1981–2010 mean. The average maximum for 2018 was 13.2°C and the average minimum was 5.7°C. These were 0.8°and 0.4°C above the respective averages for 1981–2010. The ten warmest UK years since records began in 1884 have all occurred since 2002. Seasonal averages for the UK for 1981-2010 are as follows: winter 3.5°C, spring 8.1°C, 15.8°C and autumn 9.7°C. Scotland's average temperature for this period was 7.8°C, compared to 9.7°C for Wales and 10.4°C for England. Between 1993 and 2006 the average Diurnal Temperature Range (DTR), which is calculated as the difference between the maximum and minimum recorded temperature in a 24-hour period, was 7.3°C, with a maximum of 20.2°C and a minimum of 0.8°C (first and third DTR quartiles were 5.1°C and 9.1°C respectively) (Zhang et al., 2018). These temperatures and those highlighted in the above paragraphs are summarised in figure 2.2. The goal of the following sections is to expand on this overall picture and understand how these temperatures relate to health.

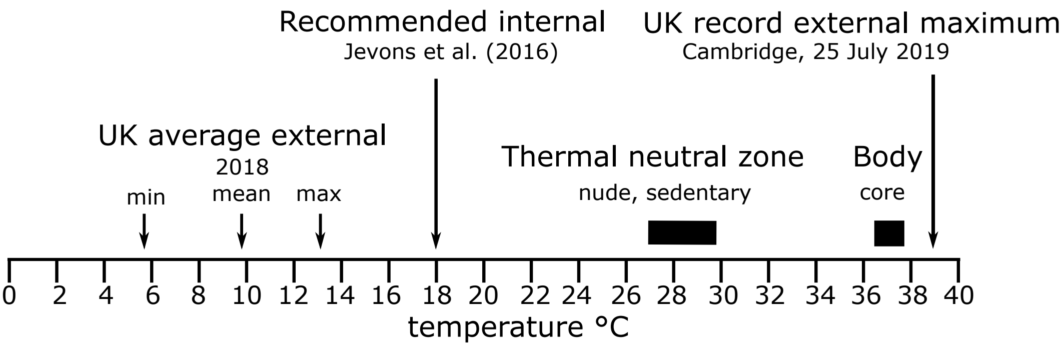


Figure 2.2: The key temperatures highlighted in the text. On average, UK internal temperatures are heated to several degrees higher than external temperatures but several degrees lower than the thermal neutral zone. This is addressed further in section 2.2

2.1 Epidemiology of cold exposure and health

The field of the epidemiology of cold exposure and health is vast and has been under development for at least 150 years. In order to render it comprehensible, some notable early papers are presented first to provide a historical overview. Following this, an examination of the current state of the literature using a number of key review papers is given. This incorporates a close analysis of the review paper 'The Health Impacts of Cold Homes and Fuel Poverty' (Marmot et al., 2011), which is frequently cited but has particular citation limitations. The majority of the papers addressed in the following sections were published during the 21st century.

2.1.1 History

The observation that ambient temperature is associated with mortality is not a new one. Macfarlane (1977) traces Western medicine's concern with the health impacts of short-term environmental changes back to Hippocrates around 2,400 years ago. Approaches more recognisable to contemporary epidemiologists emerged in the mid–19th century - Moore's 1869 study, for example, related the seasonal variation of mean temperature to mortality in Dublin, noting that "In Winter a fall of mean temperature below the average swells the numbers of cases of thoracic disease, and increases the mortality therefrom". Despite having no knowledge of the underlying causes of the diseases he was studying, he concludes that the "classes of the community on which climatic variations exercise their baneful influence in the highest degree" are "the very young, the weakly, and the aged" - this observation, stated in more contemporary language, largely holds true today.

By the beginning of the 1960's, epidemiologists were distinguishing between the impact on mortality of severe cold versus epidemics of influenza. The 1963 British Medical Journal reported no rise in winter mortality following a record breaking period of persistent cold, with a lack of influenza given as the explanation.

Rose's 1966 study into the relationship between cold weather and ischaemic heart disease is notable for a number of reasons. The paper presents evidence that there is a strong correlation between low temperatures and deaths from arteriosclerotic heart disease. The author uses the phrase 'winter excess deaths', a slight variant on the now ubiquitous term (see section 2.1.3). Furthermore, the author suggests the possibility that deaths due to cold might be subject to a delay. This observation has been made in many contemporary studies, as will be seen below in 2.1.2.

Macfarlane (1977) conducted a study into the impacts of air pollution following the 1956 Clean Air Act. At this time, the computational power available to most researchers was limited, indeed, the analysis presented is based on the visual inspection of time-series. This limitation is acknowledged by the author, but they draw a number of conclusions despite this. They suggest that influenza outbreaks are often preceded by periods of very low temperatures, and that deaths due to respiratory conditions subsequently increase. Therefore, early literature had already identified many of the primary drivers of the observed variation in mortality. Subsequent developments, which are discussed in the following section, mainly improved the mathematical models which describe these relationships.

2.1.2 Review articles

This section describes a number of contemporary review articles, published since 2012, that examine the relationships between cold exposure and health. A search was conducted using `Scopus.com`, with the criterion of "cold exposure AND health AND review" on abstracts and titles. 159 results were returned across a wide range of topics such as medicine, biology and agricultural sciences. The reviews included in this section were selected from these results based on their relevance to the aims of the review as a whole. Specific condition reviews were not included. For example, a review into hypothermia caused by antipsychotic medicine was returned by the search but was deemed too specific for inclusion.

Hajat (2017) conducted a review of the UK cold-exposure literature which sought to understand potential of future milder winters, which models of climate change project to occur. The paper sets out several key ideas and observations that are important to this study. The central epidemiological evidence comes from time-series regression and case-crossover studies which establish a U or V shaped relationship between ambient temperature and negative health outcomes - that is, there are peaks in mortality and morbidity rates at both low and high ambient temperatures.

In terms of morbidity, Hajat cites three papers, the first of which is by McGregor et al. (1999). The method uses a combination of meteorological factors, including air temperature, along with PM_{10} pollution measures (particulates smaller than 10 micrometres in diameter) to define six "air mass types". The meteorological factors were used in a principle component analysis to distinguish the different air mass types. Temperature was included with moisture levels to produce a 'hygrothermal' component. Days with similar air mass types were then grouped using a clustering technique. Three of the six air mass types had similar temperature means and ranges, but were combined with different values of the other meteorological components to distinguish them. These different air mass types and the clustered groups of days were then used as the basis of an analysis of variance test for the association with hospital admission rates for respiratory conditions on the respective days with a particular air mass type. While significant associations were found, the effect of low temperatures is not distinguished per se, since the air mass types comprise combinations of components. This makes generalisation to other regions or countries difficult. Principle component analysis and cluster analysis are subject to a degree of arbitrariness with respect to the number of principle components selected and the particular mechanisms of clustering (Agarwal et al., 2011). While the authors acknowledge these difficulties, it does mean that the results are potentially sensitive to the particular clustering method that the study used.

In the second paper, Hajat and Haines (2002) focus on low temperatures specifically and their impact on GP consultations among elderly Londoners. Once air pollution (SO_2, O_3, PM_{10} and pollen levels) were accounted for, the regression model found a 10.5% [7.6 – 13.4] increase in GP consultations relating to respiratory conditions for every degree drop in average external temperature. They found that this effect was delayed up to 15 days after the cold period. The paper suggests that it includes 'all potential confounders' in the model. While it does use an extensive list of potential confounders, it is not possible to be certain that unknown or unmeasured confounders do not influence the observed relationship. Indeed, the paper does not mention $PM_{2.5}$ levels at all, despite its known impacts on the respiratory system (Xing et al., 2016). The paper suggests that the reason for the observed associations might be due to reduced indoor ventilation or increased crowding during cold periods. This is important as it points out that some of the health impacts observed in epidemiological studies might not be due to direct cold exposure. The statistical methods used in the paper mark a step towards the more sophisticated techniques developed later by Gasparrini et al. (2010).

Finally, the study by Bhaskaran et al. (2010) also uses a smoothed regression model which incorporates a time lag to find that each degree decrease in mean daily temperature is associated with a 2% [1.1 – 2.9] increase in myocardial infarction (heart attack) hospital admissions, with the greatest effect occurring at 2–7 and 8–14 days after the cold event. Whilst the confounding effect of PM_{10} concentration is accounted for, the authors point out that $PM_{2.5}$ data were not available for inclusion in their model. This underscores the difficulty of accounting for all potential confounders. Moreover, indoor air pollutants (e.g. allergens (Carrer et al., 2001)) are likely to impact health and are not measured in this study.

Hajat (2017) found no consensus in the literature as to the impact of deprivation on cold-related health, citing a review by Tanner et al. (2013), and other studies which failed to find an effect. The author suggests that one reason for this may be the relatively higher quality of social housing in the UK, and the observation that they tend to be warmer than other tenure types (Wilkinson et al., 2001; Kelly et al., 2013) (see section 2.2 below). Hajat does not discount the negative impacts of fuel poverty on mental health highlighted by the Hill's review (Hills, 2012). While this is not a direct impact of cold-exposure, knock-on effects of the financial pressure exerted by trying to afford warmth are harmful to mental health.

Ryti et al. (2016) produced a meta-analysis and systematic review of the association between cold spells and negative health outcomes. Summary estimates derived from the papers included in the study, which analysed data from a range of global locations, were that cold spells and mortality from all or non-accidental causes have a positive risk ratio (RR) of 1.10 [1.04 – 1.17]. The specific definition of a cold spell differs across the papers included. The authors conclude that the despite the increased risk, the 'relevant patterns of exposure and induction periods remain unclear'. Furthermore, the findings do not point to any specific interventions that could aid at-risk populations. Producing meta-estimates of phenomena that have heterogeneous definitions is potentially problematic, as the estimates will likely be dominated by the studies which found the largest effects. There is also the possibility that studies which do not find an effect do not get published, a phenomena known as reporting bias (McGauran et al., 2010).

Cheng et al. (2014) sought to understand the literature on the health impacts of Diurnal Temperature Range (DTR). The review included 25 studies, from which the authors concluded that high diurnal temperature range negatively impacts health, with elderly people and children particularly susceptible. They found that cardiovascular and respiratory disease are the leading health impacts associated with DTR. They identified several knowledge gaps evident from the review. Most studies were conducted in Asia, which they suggest limits generalisability. Also, studies have not identified specific thresholds of dangerous diurnal temperature variation. They point to a study by Luo et al. (2013) which found that the immediate impacts of very low DTR – i.e. more static temperatures – are worse for health than very high diurnal temperature variation. They go on to suggest that the current understanding of confounding effects is limited, and that more work is needed on developing measures to prevent the worst impacts of DTR. A major problem with studies on DTR is establishing the extent to which the effects of the DTR variable are independent from other measures such as the maximum and minimum temperatures.

Deschenes (2014) approaches the problem of understanding the health impacts of extreme temperature exposure from the perspective of economics, in the context of improving the accuracy of climate change projections, and specifically with climate change adaptation in mind. The author finds that the majority of reviewed articles suggest that there are deleterious effects on health at extreme ambient temperature. Of greatest relevance to this study is the observation that the function that relates mortality and temperature is likely nonlinear and potentially confounded by omitted variables, seasonality and long-term trends. They also point to the diversity of exposure metrics used in different studies. For example, most public health studies used spline methods, where trends are smoothed using a weighted regression, whereas approaches from economics typically use temperature binning to differentiate varying impacts on health at different temperatures.

As mentioned above, Tanner et al. (2013) looked specifically at the impact of socioeconomic status on the relationship between cold exposure and health. Limiting their review to papers published in English between 2001 and 2011, the authors included 33 studies in total. A wide range of measures of socioeconomic status were used in the studies considered. The effects of cold on mortality and morbidity were strongest in low income households, thermally inefficient housing and households in fuel poverty. Several studies which used different measures of deprivation, such as the Townsend deprivation index, found non-significant relationships between cold and mortality. The authors suggest that future studies should examine the effects of interactions between socioeconomic and behavioural factors. Ye Xiaofang et al. (2012) conducted a review into the impact of ambient temperature on morbidity. 40 studies published before 2010 were included in the review. In agreement with Hajat (2017), several studies reported the non-linear U or V relationship between ambient temperature and morbidity. They found the threshold temperature at which negative impacts begin varies with location, as was reported in a later comprehensive work by Gasparrini et al. (2015), discussed below. They found that the majority of studies found the negative impacts of cold are delayed by up to 2 to 3 weeks. This contrasts studies which find the impacts of heat are more immediate (Sun et al., 2019; Gasparrini and Leone, 2014; Moghadamnia et al., 2017).

The review by Xu et al. (2012) focused on the health impacts of ambient temperature on children. After synthesising the findings of 33 papers, the authors conclude that extreme temperatures are associated with an increase in infectious and allergic diseases in children. They argue that children are more sensitive to hot and cold due to physiological, metabolic and behavioural characteristics. They also point to the need for more studies into the difference of temperature thresholds of the onset of morbidities between children and adults. This study is of limited direct utility to the present study, but its inclusion aims to highlight that different portions of the population can have increased sensitivity to the effects of cold.

Gasparrini et al. (2015) conducted a multi-country study of the impact of temperature on mortality using a database of 74 million deaths from 13 countries in 384 different locations. They made use of the 'distributed lag non-linear model', which is now a commonly used method of determining the risk of mortality associated with external temperature, developed by Gasparrini et al. in 2010. It is based on a non-linear Poisson distribution model of the exposure-response relationship, but aggregates the estimation across different lags which aim to capture the delayed impact of exposure. They report that cold related deaths are almost 20 times more common than heat related deaths in their study, although the results are not globally representative. Moreover, they found that the vast majority of deaths were attributable to so called non-optimal conditions, as opposed to extreme temperatures. This suggests that a small deviation away from optimal conditions leads to an increase in mortality. Overall they found that 7.71% [7.43–7.91] of mortality could be attributed to non-optimal temperature, however this estimate ranges between countries – in Thailand it was 3.37% and China 11%. For the 10 UK regions considered, the average effect was that 8.48% [7.72 –9.25] of deaths was attributable to cold, and only 0.30% [0.25 – 0.36] to heat. For London, they found the optimal temperature at which minimum temperature related mortality occurs was 19.5°C, which contrasts estimates for Wales at 16.5°C. It is unclear what contributes to this difference. The risk ratio for mortality in London increases linearly from the minimum up to a risk ratio of around 1.2 at an external temperature of 4°C. Below 4°C the risk ratio increases sharply for each degree drop in external temperature, up to a value of around 2.0 at external temperatures around freezing. The model does not account for the impact of socioeconomic status, age, or air pollutants. Furthermore, the response function between exposure and outcome is modelled as a Poisson process. As the first author points out in an earlier paper (Gasparrini and Leone, 2014), there is some controversy around the selection of response function. It is unclear how much of an impact this has on the estimates of risk of mortality.

This portion of the literature review has shown that there is clear evidence from multiple studies for the seasonal variation in mortality and morbidity. The winter peak in both mortality and morbidity is strongly associated with low external temperature. Recent developments in the statistical models used to understand these associations have showed that mortality increases with any deviation from an optimal external temperature, i.e. there is an increase in mortality associated with warmer summers as well as cold winters. The optimum external temperature at which minimum harm occurs varies with country and even city. There is evidence that the effects of the cold (and to a lesser extent warm) are delayed, from anyway between a few days to two weeks. The greatest source of uncertainty in these studies is how to account for co-variates and confounding variables, such as air pollution, which may also contribute to the observed seasonal effects. As is common with epidemiological studies, there is little consideration of the exact circumstances of cold exposure. Before considering the methods used for measures of Excess Winter Deaths in more detail, the next subsection critically reviews a report into the health impacts of cold homes.

2.1.2.1 Marmot Review: a critical analysis

In addition to the above reviews, the Marmot Review into The Health Impacts of Cold Homes and Fuel Poverty is worth particular attention. It was commissioned by Friends of the Earth and published in May 2011 to review the existing literature on the evidence of both the direct and indirect health impacts of fuel poverty and cold homes. As of August 2019, it has been cited 101 times, according to `Scopus.com`.

Unlike a traditional or comprehensive review, this section takes a different approach by examining evidence chains, namely the chains of academic references which are used to support the claims made by the paper. The aim here is not necessarily to critique the substantive conclusions of the report under examination, but instead to determine the strength of the evidence behind them. This not only helps identify the most important studies and methods, but also acts as a means of identifying potential avenues for future investigations. This approach has been inspired by several papers. First, Rekdal's 2014 examination of academic urban legends in which he describes the surprising development and evolution of the broad consensus regarding the nutritional content of spinach, which was driven by a succession of poor citation practices. He advises that [a]ccurate, complete, and relevant references to reliable sources are the best tools in order to avoid such a scenario' (Rekdal, 2014). Second, the task of tracing the origin of World Health Organisation guidelines on minimum room temperature recommendations was undertaken by Ormandy and Ezratty (2012). They note that recommendations are often given without reference to an original source, and while they suggest the "guidance was based on evidence and has been supported by subsequent research", they were not able to determine why the recommendation was changed for 15-25°C to 18–24°C between the 1960s and 1980s. Finally, Greenberg (2009) takes a related approach in creating a *citation network* to analyse the basis of a specific claim in the literature on Alzheimer's disease. The author found that biases existed because papers refuting the main claim were cited less than those supporting it. Even though the approach taken here is narrower in scope, it demonstrates a potential vulnerability of review papers.

Here, an evidence chain is the series of academic references which a claim relies on. Chains may extend to two or more papers, creating a genealogy which can extend back many years and through multiple disciplines. Once the referenced evidence is identified the process is repeated for subsequent reference levels and the content of the original claim is traced back in the literature until its source is found. A note is made of the main type of evidence that is used to support the claim at its origin, whether it be experiment, survey, interview, expert opinion or of some other type. Finally, the level of reference is noted, be it primary, secondary or a higher level.

For reasons of time manageability, the analysis was restricted to a single chapter of the review. The chapter under investigation was decomposed into 49 claims regarding health in cold homes, covering mortality (14 claims) and morbidity (35 claims). The heterogeneity of claims makes overall summary difficult. Broadly speaking, 29 of the 49 claims are supported by their references in a largely unproblematic way - i.e. the claim referred straightforwardly to the study results. The review articles in this section described above have largely been of this character. The remaining 20 claims, nearly half the chapter, rely on documents which do not provide direct references in the traditional academic sense. For example, three claims are supported by a Department of Health document from 2007 "Health and winter warmth: reducing health inequalities", and a further two claims by "Building regulation health and safety" Raw et al. (2001). This is a general problem with governmental reports; many of the claims they include appear to emerge fully formed because they are not directly cited, and they rely on the expertise of the author to sustain their authority. This makes tracing claims further extremely difficult.

For 7 out of the 49 claims it was difficult to determine whether the cited documents support them fully. This was due to either referencing mistakes or the claim relating to a report which was referred to in general terms. One such example is instructive. The claim that EWDs are not related to socioeconomic deprivation uses four sources to support it. However, three of these sources do not discuss deprivation specifically, and the fourth suggested the opposite finding (Healy, 2003). Of course, this is not to say that the claim is necessarily false. Indeed, the paper by Hajat (2017) given above reached a similar conclusion. However, it does highlight a particular difficulty with review articles that need to synthesise large amounts of contradictory or partly incommensurate findings. This is not a problem of the Marmot Review alone, but one that effects academia more broadly, as is reflected in current concerns about the replication crisis in energy research (Huebner et al., 2017). It is also important to highlight that the authors of the Marmot Review not be singled out, the approach adopted here could have easily been applied to any of the reports which sit at the intersection of academia and policy.

The approach adopted in this section is not without limitations. It is time intensive, and most academics simply do not have time to check every citation in close detail, let alone second order citations of citations. A large part of the responsibility of the negative impacts of poor citation practice and the propagation of unsupported evidence must lie at the point of peer-review, which in the specific instance of the Marmot Review did not take place, since it was a commissioned report. Due to their informal citation structure, reports conducted or commissioned by non-academic bodies arguably carry the greatest risk of disseminating unverifiable claims. Notably, this approach did not uncover any information which challenges the broad, substantive conclusions of the Marmot Review, such as the observation that more people die in winter than in summer in the UK. However, producing overarching narratives across heterogeneous evidence sources, especially when citation and reference practices are in places poor, risks occluding both uncertainties and the detail of the specific ways that cold temperatures negatively impact health. This is of particular importance for this thesis in forming specific understanding of the origin of the observed relationship between cold exposure and health.

Taking the executive summary as an indicator of the overall findings of the review suggests the report relies on the findings of a small number of studies to provide general conclusions. This is especially the case with regards to the findings on excess winter deaths (EWDs). The following section reviews some of the important papers on EWD and shows that it is a useful concept, although not one without problematic methodological aspects.

2.1.3 Excess Winter Deaths (EWD)

The concept of EWD was introduced in the first chapter of this thesis. The EWD characterises the number of deaths in winter over the average at other times of the year. It remains widely used in cold-exposure related health studies. This section reviews key papers relating specifically to EWD, since it is key to understanding the dynamics of cold related mortality.

Wilkinson et al. (2001) sought to understand the determiners of EWDs using two linked datasets, one providing mortality data for England between 1986 and 1996, and the English House Conditions Survey (EHCS) to provide housing and demographic data. The regression model used to link these data found few significant results. The exception was the age variable, for which the relative risk of dying in winter increased to 1.28 [1.13 – 1.46] for those aged 85+, compared to those aged 0–44. There was some suggestion that the relative risk of mortality was lower in newer houses, however this trend was absent once age, gender, socioeconomic status and modelled indoor temperature were accounted for. This is an example of where it is important to distinguish between association, prediction and causation. As Huitfeldt (2016) points out, most epidemiological studies are very careful to avoid causal claims – but they are often imprecise as to their specific meaning of the term risk factor. Huitfeldt (2016) argues that whether or not a model controls for other factors should depend on the research question that the authors hope to address. In the case of the Wilkinson et al. (2001) paper, living in a new home is associated with a reduced risk of EWD, but if one wishes to predict such a status, then the factors (age, gender etc) are better at doing so. The paper, and many of the other papers reviewed here, would benefit from making this distinction clear.

Using temperatures lagged by 3 days recorded between 1970 and 1999, and mortality data, Donaldson and Keatinge (2002) calculated that 2.4% [2.0 – 2.7] of EWDs could be attributed to influenza. The method of attributing cause of death likely creates the largest uncertainty in their estimate. They conclude that the remaining winter deaths must be attributable to cold exposure. However, the paper does not discuss any other seasonal mechanisms that might contribute to the observed winter peaks, such as indoor or outdoor pollution levels. They conclude by emphasising the importance of outdoor cold exposure, and as with all of the studies looked at in this section there is little discussion of exactly how cold exposure occurs.

Healy's 2003 paper comparing winter mortality in 14 European countries has been cited over 550 times according to Google Scholar. Winter was defined as the four months December to March, and a regression model comparing the relative winter death rate against a number of variables which characterise health in each country. The model found the highest rates of seasonal mortality occurred in the mildest countries. This apparent paradox, the author argues, is resolved by the model's indication that winter related mortality was found to be highest in the countries with housing of the poorest thermal efficiency - namely Portugal, Ireland, Greece and the UK. High levels of poverty, income inequality, deprivation and fuel poverty status were also found to be significant predictors of the winter mortality rate.

Hajat and Gasparrini (2016) report a number of methodological issues with the calculation of EWDs in the simple manner described in section 1.3 and above. They point out that the cold period in England, for example, is concentrated in the four months December to March, whereas Scandinavian countries have a longer duration cold period. This makes EWD a largely unsuitable comparison between countries in the manner of Healy's paper mentioned above. Furthermore, they argue, many cold related deaths fall outside the winter period. The measure can also be heavily influenced by a high summer mortality, which would reduce the apparent number of excess winter deaths. The authors then compare the simple EWD approach to their own distributed lag nonlinear model, which is the same method mentioned above in the work by Gasparrini et al. (2010). They suggest that the use of EWDs led Staddon et al. (2014) to incorrectly conclude that climate change will not reduce winter related deaths – the use of the distributed lag nonlinear model would find a strong cold weather effect that will remain even as the climate warms. A later paper by Gasparrini et al. (2017) examined the effects of future climates on excess winter mortality across a range of global regions. For Europe, it showed a moderate reduction in cold related mortality and a sharp increase in heat related mortality as a function of increasing emissions, thereby challenging the findings of the Staddon et al. (2014) paper. However, long-term projections are inherently uncertain and fall outside the scope of this thesis.

A review of different methods for calculating excess winter mortality is given by Liddell et al. (2016). Alongside the standard EWD method, they outline a method using heating degree days, which accounts for heating demand, rather than using a fixed winter period to asses deaths. They observe that only 50% of heating demand in Ireland, for example, occurs December to March - indeed, for only 2 of 30 European nations does the 4-month method adequately reflect heating degree days. Overall, despite the convenience of the EWD method, it is largely ineffective at quantifying the health impacts of cold in comparisons between countries. This finding agrees with the results of the paper by Hajat and Gasparrini (2016) above. Overall, it is apparent from the literature that the EWD measure provides a simple and useful means of characterising cold-related mortality, but that it is no appropriate for comparison across different climates. This applies for both differences between countries with different climates and for future climates within the same country.

2.1.4 Summary

This concludes the first portion of the literature review, examining the large scale features of the relationship between cold and health. It corresponds to the leftmost image in the summary figure 2.1. The first sub-section examined the notable literature reviews related to the epidemiology of cold-exposure. The overall picture is one of clear evidence relating cold to increases in mortality in morbidity, but, as might be expected from epidemiological studies, detailed insight into the specific ways that harmful exposure occurs is not addressed by these studies. There is good evidence to suggest those with health conditions are most adversely impacted by cold exposure. Older people are also more vulnerable to the impacts of cold.

There is a ubiquitous issue of how, and indeed whether, to account for confounding factors such as air pollution and sociodemographic status. The systematic reviews are likely impacted by issues such as publication bias. A close examination of the citation structure of the Marmot Review revealed numerous citation errors. The field as a whole would benefit from improved methods to account for bias when constructing systematic reviews. The overview of the EWD literature also revealed issues, mainly relating to the difficulty of generalising EWD across different winter durations. However, the metric itself is useful and easily calculable for analyses within particular climates.

This portion of the review provides the backdrop for the thesis as a whole. The next section looks at the next level of the relationship between cold exposure and health, namely that which occurs at the level of the household and the individual.

2.2 The temperature of UK homes

The previous section gave an overview of the development and contemporary state of the epidemiological literature relating cold and health. The large majority of these papers do little to consider specifically where and when harmful cold-exposure occurs. For example, there is a significant open question as to the relative importance of outdoor versus indoor cold exposure (Keatinge and Donaldson, 1997). The following section examines the evidence at the domestic level with a review of what is known about temperature distributions in homes, in order to better understand the differences in cold exposure as a function of demographics and housing factors. This is important given that adults spend an average of 52% of their time at home (Daraktchieva, 2018). Significant differences as a function of demographic or building factors may contribute to the differences in observed temperature related mortality and morbidity rates described in the previous chapters. This is particularly important given this study aims to understand how exposure to cold varies at the individual level. However, there are also differences in the physiological responses to cold as a function of demographics. The section which follows this is therefore devoted to giving an overview of thermophysiology. Taken together, these sections correspond to the central portion of figure 2.1.

2.2.1 Temperature patterns in homes

The papers considered here have been collected throughout the course of the PhD. They resulted from searches relating to indoor and domestic temperatures. This section is not designed to provide a systematic review of the literature, rather it focuses on recent key papers, as well as some earlier studies, which help to shed light on the current state of knowledge of domestic temperatures in the UK.

A review by Vadodaria et al. (2014) of UK domestic temperatures provides the history of field measurements between 1969 and 2010. The article references the government report of 1961, "Homes for today and tomorrow" which recommended that heating systems in homes be able to achieve temperatures of at least 55°F (12.8°C) in kitchens and hallways, and 65°F (18.3°C) in living areas (Morris, 1961). This contrasts the whole-home minimum temperature recommendation of 18°C of Jevons et al. (2016) described in the opening section. Based on their review of small scale domestic temperature studies conducted in the 1960s and 1970s and larger scale studies conducted more recently, Vadodaria et. al conclude that there has been little or no increase in the temperature of occupied living areas. Bedrooms, on the other hand, have probably increased in temperature during winter and spring-time between 1969 and 2010. They suggest that these changes are most likely attributable to the increased prevalence of central heating in British homes.

As highlighted above in the section on EWDs, the widely cited paper by Wilkinson et al. (2001) made use of survey data and predicted hall temperatures from 1986 to 1996 to produce an estimate that the odds ratio (see section 4.7.2.3) for winter vs summer mortality in the coldest 25% of homes is 1.5, compared to 1.3 for the warmest 25% of homes. However, the validity of the temperature predictions is difficult to determine because they are not compared to measured values, and so should be treated with a degree of caution.

Oreszczyn et al. (2006) report standardised temperature readings in English homes, recorded during the Warm Front home insulation improvement scheme in winter 2001/02 and 2002/03 in Southampton, Newcastle, Manchester, Liverpool and Birmingham. Standardisation is necessary to compare readings taken at different external temperatures. Across the whole sample, the median standardised daytime living room temperature was

19.1°C, and the median standardised night time bedroom temperature 17.1°C. The study found homes that were older were colder, as were those that had lower insulation levels, younger inhabitants, and greater difficulty paying bills. An analysis of the Warm Front scheme (Green and Gilbertson, 2008) found the installation of home energy efficiency measures increased living room temperatures by 1.6°C on average (from 17.9°C to 19.6°C) and bedroom temperatures by 2.8°C (from 15.9°C to 18.3°C). A total of 1,600 dwellings were monitored for the study. Measurements of temperature and humidity were taken at half-hourly intervals for periods of between 2 to 4 weeks. Given there are around 28 million dwellings in the UK, of which around 10% are fuel poor, and that the winter period comprises 12 weeks of the year, the total monitored hours represent a maximum of 0.2% of overall UK winter domestic hours for the target population. The authors do not comment on the extent to which the population is representative of UK fuel poor households in general. The authors do report a mismatch between who received the Warm Front interventions and who was found to be fuel poor. The problem of targeting and methods is central to the fuel poverty literature (Ambrose and Marchand, 2017), but outside the scope of this thesis. Moreover, there is no information as to how long the occupants were present in their homes during the monitoring period, either pre or post intervention. This means it is difficult to determine the extent to which the occupants were exposed to the conditions within the homes. The authors acknowledge that the absence of an effect as a function of socioeconomic status may be due to the fact that the index of multiple deprivation is only available at a relatively course scale which would not be able to distinguish differences in the sample that is generally less well off as a whole. Counter-intuitively, while fewer people reported difficulty with paying their bills after the efficiency measures were installed, the energy use of the homes was measured to increase. This highlights a complexity associated with subjective perceptions.

Huebner et al. (2013) use living room temperature data from the Carbon Reduction in Buildings Home Energy Survey (CarB HES). The mean monitored living room temperature in winter was calculated across homes in the survey. It revealed that temperatures drop over night from around 20°C to around 18°C, and then rise again through the day as the home is heated, first in the morning and then in the evening after work. They found a significant temperature difference of 0.16°C between weekday and weekend temperatures. The substantive conclusion of the study is that the level and duration of heat demand assumed in common UK housing stock models (specifically, the Building Research Establishment Domestic Energy Model (BREDEM)). The paper would have benefited from an analysis of what the differences between modelled and observed heat demand would mean for national domestic heating demand.

Huebner et al. (2015) followed up this study by conducting a cluster analysis on the same data. This approach identifies four main home temperature profiles - a significant relationship between the ages of those present in the household and each of the temperature profiles, as did the building and main heating type. They also found that while one of the clusters matched the expectations of building models such as BREDEM, the other 3 made up over half of homes - which may account for a substantial amount of the performance gap of such simulations. As with the study by McGregor et al. (1999) discussed above, clustering techniques are potentially sensitive to the particular choice of clustering algorithm which defines the study groups. Taken together, these papers underscore the heterogeneity in internal temperatures, which may or may not correspond to demographic and building factors. As with all the models described here, unexplained variance is a significant factor in the actual temperatures of homes.

Kelly et al. (2013) also made use of CarB HES data, but sought to predict whole-year indoor temperatures using a novel panel method. The model used was able to predict 45% of the variation in internal temperatures. They calculated that for a mean internal temperature of 19.6°C the external temperature and geographic location accounted for approximately 6.8°C of the variance. Behavioural variables may explain up to 2.9°C and sociodemographic factors may explain up to a maximum of 3.7°C. Of this, they found that tenure accounts for up to approximately 1.4°C of the variation in internal temperatures. They also found that that tenure also has an impact on internal temperatures. Relative to owner occupied dwellings, housing association homes were 0.49°C warmer, rented dwellings 0.94°C warmer and council houses 1.37°C warmer. As with other studies in this section, no consideration has been made as to the extent to which occupants are exposed to the observed differences in indoor domestic temperatures. On the basis of the estimates of home occupancy time given at the beginning of this review, one might expect that those who stay at home would be more impacted by the average temperature of a dwelling than those who have full–time jobs.

Mavrogianni et al. (2013) surveyed contemporary and historical data sources to trace the evolution of average UK indoor temperatures over the last decades. They found, subject to a good deal of uncertainty surrounding methodological differences, an increase in mean internal temperature of up to 1.38°C per decade between 1978 and 1996. This is consistent with the findings of the review by Vadodaria et al. (2014) described above. They also suggest that this increase, if it is correct, may account in part for increasing obesity rates in the UK, due to reduced energy expenditure on thermoregulation. They also suggest that "the study of past exposure is limited to data on room conditions. Future research should refocus from the average temperature conditions in buildings to measuring the overall personal exposure of an individual in both domestic and non-domestic environments". This finding is particularly relevant to the present study which is motivated by understanding cold-exposure of individuals. This point is also highlighted by Johnson et al. (2011) in their review of temperature homogenisation in the US and UK domestic stock. They conclude "studies of ambient temperatures in specific locations cannot capture a detailed picture of the frequency and duration of thermal exposures in the course of an individual's daily activities, and there is a need to document personal and socially patterned trends in cold exposure". A later study of over 100,000 adults in England found that BMI was inversely related to domestic temperature (Daly, 2014). Specifically, those living above 23°C were found to have lower BMI than those whose homes were measured to be cooler. This relationship persisted when accounting for covariates such as age, social class and health. However, as was suggested above, the use of spot measurement in the study cannot account for overall cold and warm exposure.

A study by Hamilton et al. (2017) addressed the determinants of living room and bedroom temperatures, as measured by the Energy Follow Up Survey (EFUS), a large cross-sectional survey of dwelling characteristics and indoor temperature measurements in England. Living room, hallway and bedroom temperatures were monitored using modified TinyTag Transit 2 data loggers every 20 minutes for 943 households. They found that standardised living room temperatures of houses built between 1945 and 1990 were significantly warmer by around 1.3°C than those built pre-1900, when standardised to external temperatures of 5°C. Local authority living room temperatures were 1.6°C warmer than owner occupied homes, but bedroom temperatures were not significantly different. Homes with double glazing were around 1.4°C warmer than those with single glazing. Homes with retired occupants had living room temperatures of 1.0°C higher than the mean, and those who were classified as vulnerable (i.e. on means tested or certain disability related benefits) were 0.6°C warmer than those who were not. The paper uses three different temperature standardisation procedures and it is not clear which one constitutes the most important model.

Huebner et al. (2018) conducted a study which also used data from the EFUS dataset. This sought to understand whether households met the recommendations of 18°C (per the recommendation of Jevons et al. (2016)). The study compared the wintertime temperatures of occupants who reported having a long-term disability (LTD) and those who did not (no LTD), as well as those aged 64 and younger (the younger group) to those who were 65 and older (the older group). It found significant differences in mean bedroom temperatures: LTD mean was 18.35°C; no LTD mean 17.87°C (difference 0.48 [0.03 – 0.94]°C). Significant differences in living room temperature were found for both LTD status and age. Temperatures were higher in the LTD group (19.37°C) versus the no LTD group (18.50°C) difference 0.87 [0.44 – 1.30]°C. The older group had warmer mean living room temperatures (19.32°C) than the younger group (18.55°C), and the difference was

0.77 [0.34 – 1.21]°C. Risk ratio estimates for the likelihood of having temperatures at or above 18°C were also given. It found that people with LTD are 1.30 times more likely to have dwelling temperatures which met the recommendation compared to the no LTD group. The older group was found to be 1.56 times more likely than those below 65 years to meet the 18°C recommendation. The study did not report the confidence intervals on this risk ratio estimate. The authors acknowledge that occupancy is difficult to assess for these groups – it is plausible that participants without an LTD, and those who are younger, were at home for less time than the comparison groups. This paper is useful for the present study as it provides information regarding the different measured dwelling temperatures as a function of the health of the occupants. Whether those who have LTD chose warmer temperatures in order to alleviate the symptoms of their conditions, or because they are less active and have thermal comfort preferences which lead to higher temperatures, is not determinable from the results.

As a whole, the papers reviewed suggest a deal of heterogeneity across UK domestic temperatures. Predominantly, homes have been getting warmer on average over the last 50 years, but mainly due to an increase in bedroom temperatures. There is evidence to suggest that old homes, particularly those built before 1900, tend to be colder than those built more recently. Local authority housing tends to be warmer than privately owned and rented housing. The next section considers the thermophysiological response to cold, specifically to understand whether different demographic characteristics correspond to differences in responses to cold exposure.

2.3 Thermophysiology

The majority of thermophysiology studies involve highly controlled experiments of small numbers of participants. This approach is radically different from that of epidemiology, and trades off generalisability for accuracy, precision and a stronger adherence to the scientific method. The field as a whole can be considered an enquiry into the physiological mechanisms which underpin thermoregulation. This entails understanding thermal sensation and the determiners of thermal comfort, as well as circadian rhythms and longer term adaptation mechanisms (Hensel and Schafer, 1984). Early studies such as Keatinge et al. (1964) measured the physiological responses to cold water showers, finding large increases in blood pressure and heart rate following exposure to an ice shower (with a water temperature measured to be between 0.0 and 2.5°C) for 2 minutes at 6 litre/minute. Later studies examined the specific response of physiological systems to cold exposure, such as Rutkove (2001) who describes the neurological response and potential for damage to extreme cold and heat exposure. Respiratory conditions, such as chronic obstructive pulmonary disease (COPD) are understood to be exacerbated by cold weather

(Serra-Picamal et al., 2018). However, the capacity of the human body to adapt to cold exposure is also notable - a review by Daanen and Lichtenbelt (2016) describes a number of cold adaptive mechanisms which can differ as a function of intensity and duration, and with age, gender, exercise, diet and ethnicity. The effect of repeated life-time cold exposure is less clear. Cold water immersion can lead to reduced metabolism, but cold air exposure may increase it. Some degree of insulative adaptation may occur, but it is difficult to discount the impact of higher calorie diets in cold regions. Furthermore, van Lichtenbelt et al. (2017) report that mild cold exposure, outside the temperatures at which people are thermally neutral, can promote glucose metabolism – 10 days of intermittent mild cold exposure was found to increase insulin sensitivity in patients with type 2 diabetes. The authors also suggest the cardiovascular system of healthy subjects may be positively affected by regular exposure to cold and heat. The neurophysiology of skin thermal sensation is described in detail by Filingeri (2016). While a detailed consideration of these mechanisms is outside the scope of this study, it is helpful to keep in mind some of the conclusions of the paper; humans' ability to detect variations in humidity and temperature both within our bodies and in the external environment entirely determines our thermoregulatory behaviour. There is evidence that thermosensory mechanisms are 'intrinsically intertwined' with the central processing of pain (Filingeri, 2016). The authors suggest that understanding this facet of human physiology may lead to novel pain treatments. Overall, the papers reviewed here suggest that whether or not someone has morbidities is key in determining the effectiveness of metabolic responses to cold. This observation is an essential aspect of understanding the relationship between cold and health. It is contrasted by emerging evidence of the specifics regarding thermal adaptation. People who are acclimatised to cold exposure may have a diminished risk of cold related strain. This is important as it suggests that the particular response of exposure to, for example, an ambient temperature of 12°C (which is far from the thermal neutral zone) depends on the personal history of the individual, as well as whether or not they have morbidities. In an editorial for the journal Experimental Physiology, Tipton (2019) uses a mixture of evidence from multiple studies to take this point even further. He hypothesises that with regards to humans being healthy, "the dynamic equilibrium that underpins homeostasis needs to be perturbed", that is to say, good health is intrinsically linked with being exposed to thermally varying environments.

This concludes the central portion of the literature review. The next section is concerned with the final level of the review – the right most portion of figure 2.1 – which relates to the variation of wrist temperatures between individuals, and attempts which have been made to measure the local environment of the individual by using sensors worn on the body. The latter topic is addressed first. The former is considered using a comprehensive review. This is to ensure that all available evidence in the literature regarding the variation in wrist temperature is available. From this, the overall picture of the literature is combined into a conceptual model. This, and the development of the thesis research questions, will be the focus of the next chapter.

2.4 Experienced temperature

There is presently no research which seeks to characterise or measure the personal cold exposure of individuals in real-world settings. Two studies which were carried out in warm urban environments do provide some insight. First, Kuras et al. (2015) report the results of a novel study to measure the individually experienced temperature (IET) of study participants during a Boston heatwave. The study recruited 23 participants who wore an iButton temperature sensor on a belt loop, or attached to a handbag carried at all times, for one week at a 5-minute period (which was subsequently averaged to 1-hour). Despite the adversely hot external conditions, the participants' IET were significantly cooler than the prevailing ambient external temperature, but fewer than half of the participants experienced statistically significant temperature differences between the heat-wave and reference periods. This may suggest that the participants were seeking refuge from the adverse levels of heat. Moreover, the study revealed heterogeneities in IET within the same neighbourhood. They found that older participants were more likely to have higher IET than younger participants. The authors conclude that ambient temperature may misrepresent experienced temperature. The study is limited by the lack of consideration of body heat on the sensor, the small sample size and short duration of the study. The authors do not estimate the impact of body heat on the iButton readings, but they do point out that radiant energy from sunlight might have impacted the readings. Despite these limitations, the study is innovative in its attempt to measure experienced temperature.

Second, Nakayoshi et al. (2015) sought to measure both the prevalent ambient conditions along with physiological variables in the urban environment of Tajimi city, Japan. The study recruited 26 healthy male and female subjects ranging from 23 to 74 years in age. The Standard Effective Temperature (SET) was recorded for each participant as they traversed a predetermined urban route. SET is a metric that factors in mean radiant temperature as well as relative humidity, air velocity, activity level and clothing, in order to standardise the temperature across different environments. The experiment revealed wide variations in SET in the complex urban environment. However, the technical complexity of the set-up used in the study makes its deployment at a population level impractical.

Tamura et al. (2018) gives a review of the current state of wearable thermometers. The vast majority of them are designed for medical monitoring of body temperature. Aside

Exposure type	Description	Representative physiological processes	Example
Average	Arithmetic or geometric mean of past exposures	Slowly or partially reversible effects	Pulmonary irritants
Cumulative	Product of intensity and duration	Cumulative, irreversible effects	Silica and silicosis
Duration	Start of exposure to onset of disease	Cumulative, irreversible effects	High level noise exposure and hearing loss
Peak	Various measures of short term, high exposure periods	Reversible, inflammatory processes	Strain on lower back and back pain

Table 2.2: Common summary measures of exposure (reproduced from Kriebel et al. (2007))

from two studies mentioned, no other studies report the use of monitors which have been developed with the express task of characterising the immediate thermal environment of the individual. The next section considers the potential impact of wrist temperature variation on a wrist worn monitor such as the AX3.

A final important consideration of this section of the review comes from the field of exposure and dose modelling relating to occupational health. Although this field does not consider cold exposure to be equivalent to chemical or other pollutant exposure, it is instructive to understand how exposure and dose are considered by the field. A comprehensive outline of the various approaches to modelling exposure is given by Kriebel et al. (2007). They highlight the need for exposure to be quantified in terms of a metric prior to modelling. Mathematically, this amounts to transforming the time-varying signal of exposure into a summary which quantifies the total exposure. This exposure will in turn impact different organs to varying extents. In all but a few instances, measuring the burden on particular organs is unfeasible. Kriebel et al. (2007) highlight four common summary measures of exposure, summarised in table 2.2 below.

An explicit assumption of the metrics described in table 2.2, is that a directly proportional relationship exists between exposure and risk. As the distributed lag non-linear models described above show this is not the case for exposure to external temperatures. For more complex relationships a dosimetric approach is more appropriate. For these, the degree of response to a dose is variable over time, and can depend on an individual's previous dose history, as well as other factors. Under this scheme, the weight given to a particular level of exposure may be a non-linear function of its magnitude. It is important to keep these observations in mind when understanding experienced temperature as it relates to cold exposure. At present, Cold 'dose' is not a concept that appears in the literature.

2.5 Wrist temperature

This final section of the review seeks to understand what evidence exists in the literature for the variation of wrist temperature as a function of sociodemographics, activities, and medical conditions. This review in particular was designed to be systematic. A broad search was conducted on both `Scopus.com` and PubMed (`ncbi.nlm.nih.gov/pubmed/`) of abstracts and titles containing the words "wrist" and "temperature", which returned 1319 results. *Epii-reviewer 4* was used to screen and process the results, as shown in figure 2.3. A total of 35 papers remained in the review, and are discussed in the following sections. The papers have been grouped by theme to aid readability.

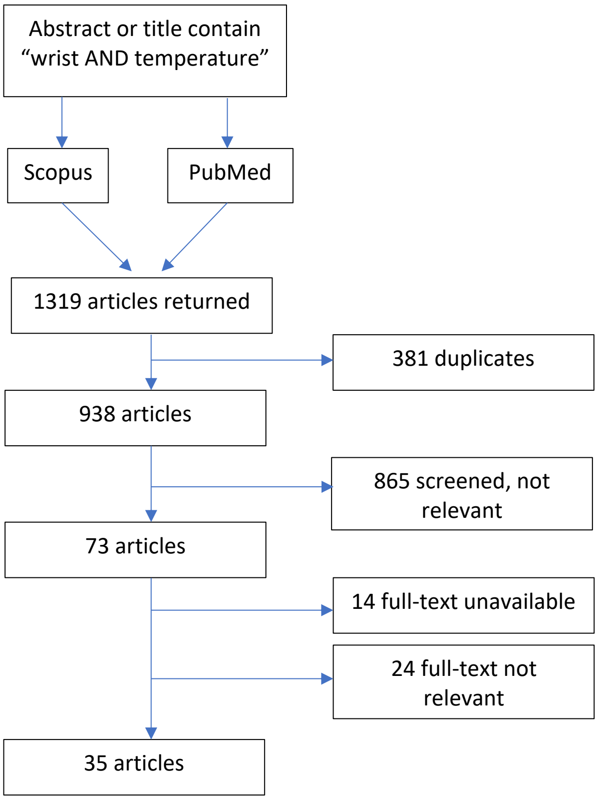


Figure 2.3: The screening process for the results returned by the literature search for 'wrist temperature'.

2.5.1 General papers

Areas et al. (2006) analysed the temporal variation of oral, axillary (armpit), thorax and wrist temperature (WT) of 12 healthy 25–year–old men and women. The analysis made use of the cosinor fitting technique (Cornelissen, 2014) for which temporal variation of a signal is fitted to cosine wave. From this cosinor analysis, the fitted wrist temperatures ranged between 32°C and 36°C, with a mesor (mean) of 33.01°C (sd: 0.14°C). The average amplitude of the WT variations was 0.79°C (sd: 0.04°C). The paper would have benefited from reporting the unfitted variation in wrist temperature too. Sarabia et al. (2008) also measured the variation with time of WT, but instead recruited 99 university students. The authors found WT positively correlated ($r = 0.9$) with the sleep periods, suggesting warmer WT is associated with sleep periods. The studies discussed here tend to be limited by a lack of consideration of the heterogeneity of populations – they focus on small groups of similar people. The impact of different seasons on wrist temperature is considered by Wey et al. (2012). For a group of 24 adult Guarani natives living without access to electricity in Brazil, winter cosinor fitted mean WTs are lower than summer ones: the summer minimum was 33.7 ± 0.7°C and the winter minimum 32.0 ± 1.2°C. The study would have greatly benefited from the explicit measurement of external temperature, and an understanding of the direct relationship between that and wrist temperature.

Several studies present evidence that wrist temperature measurements, used in different ways, can be used as a proxy for thermal sensation. Choi and Loftness (2012) report the use of the rate of change of wrist temperature as a good predictor of thermal sensation. Jacquot et al. (2014) measured the skin temperature at 14 locations on the body of 16 healthy females aged 18 to 30 under two experimental protocols, one of a gradually warming ambient temperature and the other which gradually cooled. They found that wrist temperature was the best predictor of thermal sensation. Sim et al. (2016) took measurements from 8 people in Switzerland and found that a model based on a combination of three wrist temperature measurements was the best predictor of thermal sensation. In a study of 430 office workers in China, Wu et al. (2017) also found that upper extremity skin temperatures (finger, wrist, hand and forearm) are good indicators of thermal sensation. Choi and Yeom (2017) found that using the back of the wrist temperature combined with its rate of change predicts thermal sensation vote with 93% accuracy. These studies tend to suggest that the use of wrist temperature, or its rate of change, is a good proxy for thermal sensation.

The search criteria returned a paper which considered heat flow to the hand. By using sensors placed at the wrist, Levy et al. (1977) measured the heat flow in the hand of 10 participants using a water bath. They conclude that "[h]and blood flow is extremely variable, but always exceeds the hand's metabolic needs. In effect, hand blood flow is principally determined by the thermoregulatory requirements of the organism". This paper discusses the paradoxical *hunter's reaction*, for which immersion in cold water (around 4°C) leads to a vasodilation, and thus increased blood flow to the finger tips. It is reviewed extensively by Daanen (2003) and pertains chiefly to cold water immersion and so will not be considered further in this review.

Campos et al. (2016) do not report a convincing method of distinguishing different age groups using WT measurements. They make use of a metric defined as the minimum of a moving average within a two hour window. From this they suggest they can distinguish the wrist temperature rhythms of older people from younger people, but the complexity of the features used in this paper makes generalisation to other studies very difficult. As a result it is largely unhelpful for understanding wrist temperature variations more generally.

Howell (1983) studied 103 elderly women and the temperature gradient between their axilla (armpit) and fingertips. They found that "there was a fall of skin temperature between axillae and thumbs amounting to 4.8°C. ...Skin temperature readings could be correlated with the temperature of the circumambient air but not with age." Marin et al. (2011) found a significant decrease in the phase complexity of wrist temperature with age. However, the phase complexity is an unwieldy characterisation of a time-series and there is not theoretical justification given by the authors as to why this is a useful quantity to measure. These findings are important for the present study as they might suggest that any differences recorded as a function of age by a wrist worn temperature monitor would correspond mainly to differences in ambient environmental conditions and not differences in wrist temperature.

Chilcott and Farrar (2000) found no significant difference in WT between genders in their study of 8 males and 9 females. Again, this finding is important for the present study as it might suggest that sex differences can be attributed to ambient temperature differences. However, Shilaih et al. (2017) found that ovulation impacts WT in their study of 136 women in Switzerland. After monitoring both the WT and menstrual cycle they found the average early-luteal phase WT was 0.33°C higher than in the fertile window. However, 18% of the subjects did not show such a pattern, and the authors concluded that WT variation alone does not capture ovulation events. This does not discount the possibility that during an ovulation event women might alter their environmental conditions, and thus the readings of a wrist worn temperature monitor. However, the difference of 0.33°is small, and unlikely to be detected.

Martinez-Nicolas et al. (2013) sought to understand the different exogenous factors which mask the underlying variation in WT. The study recruited 103 subjects aged $18 - 24$ and measured their WT every 10 minutes, as well as activity, body position and sleep. The cosinor fitting found that amplitude was most affected by environmental conditions. There is a potential limitation to this finding in the use of cosinor fitting, since this presupposes a functional form of the variation in wrist temperature. Despite this, the finding that the amplitude of variation of WT is most impacted by the environment means that a wrist worn sensor would be influenced in a manner which accords with ambient temperature. The counter-factual case, for which wrist temperatures increased in cold conditions and decreased in warm conditions, would have rendered a wrist worn temperature sensor less sensitive to ambient temperature.

2.5.2 Activities

The following subsection considers the evidence for the impact of specific activities on the WT. This is important for the present study as it provides information on the extent to which a wrist worn sensor would be impacted by the particular activities of the wearer, separately from the environmental conditions that they may have chosen for a given activity. This point will be synthesised in the following chapter on the conceptual model underlying this study.

Baritz et al. (2013) present evidence from thermographic pictures which indicates that hands, and to a lesser extent wrists, warm as the participants carry out hand exercises such as repetitive gripping. This might suggest that a wrist worn temperature monitor would read a higher temperature when the participant was active, although the differences that the study reveal are not quantified. Carreiro et al. (2015) report the results of a small scale pilot study of 3 participants whose bio signals (including WT) were measured in the time surrounding intravenous administration of opioids, and Carreiro et al. (2016) followed up this trial with a study of 30 men and women. An average increase of 2.62°C in WT was recorded when comparing before and after opioid administration. No significant differences as a function of gender, age or opioid type were found. It is not expected that opioid administration is sufficiently common to effect wrist temperature. A paper returned by the search criteria gives a limited description of the measurements of the WT of two participants and reports that they are not normally distributed (Camargo et al., 2012). There is no reason to expect that WT measurements should be normally distributed at the level of two participants, so this paper is of limited use to this study. Finally, Nissilae et al. (1996) report a limited experiment using two participants which found WT dropped more than thorax and axilla temperatures when a participant spent 30 minutes at very cold ambient conditions of -10°C. This finding is commensurate with the observation that vasoconstriction limits blood flow to the hands during cold exposure.

Several papers have looked at the impact of shift work on circadian rhythm measured using WT. Jang et al. (2017) report their cosinor analysis of the WT of 68 day workers and 53 night-shift workers in South Korea. They found significant differences between the WT amplitude between the groups, suggesting that shift work disrupts circadian rhythms. The amplitude of WT for the shift-workers fell as the night shift progressed (0.92 to 0.85°C) and continued to do so following a rest period (0.69°C). It rose again during the morning-shift days (0.82°C). Day workers on the other hand had higher amplitude WT (0.93°C) than those workers who had gone back onto morning shifts. These results accord with an earlier study of nurses by Ferreira et al. (2013) which also found that shift work disrupts circadian rhythms as measured using WT. Bracci et al. (2016) also measured the WT nurses working in the day and compared them to night–shift workers. Following cosinor fitting, the authors found that night–shift workers have a higher mesor, similar maximum and higher minimum, and so a decreased amplitude. This result is consistent with the other studies mentioned. Furthermore, they draw a potential link between working shifts and the development of metabolic syndrome due to the similarities in WT variation in both groups. These findings are important as they suggest that the disrupted sleep patterns that are associated with shift-work may impact temperature measurements made by a wrist worn temperature sensor.

2.5.3 Pathologies

This subsection considers the impact of pathologies on wrist temperature and its variation. Understanding the impact of different illnesses is essential when trying to understand the link between ambient temperatures (as measured by a wrist worn sensor) and health. Areas et al. (2006) highlight earlier work, outside the scope of this review, which suggests associations between variations in circadian rhythms and infections of the central nervous system and pneumonia (Cunha and Tu, 1988). Harfmann et al. (2017) conducted a study of subjects aged 50 to 70 to understand the relationship between markers of circadian rhythm (WT amplitude and stability) and measures of metabolic syndrome. They found that only triglyceride levels correlated with a low WT amplitude and stability. However, the monitoring of complex bio-markers of circadian rhythms such as triglyceride is not practical for a large scale population study.

Corbalán-Tutau et al. (2011) compared the time evolution of WT over a three-day period of 20 normal-weight women with that of 50 obese women. The WT of the obese women was found to be significantly lower (by 0.3°C) than the normal weight women. Their 24-hour WT pattern was also flatter, in a finding consistent with the Harfmann et al. (2017) study. A later study by Corbalán-Tutau et al. (2015) found that obesity and metabolic syndrome could be characterised by WT when coupled with questions about sleep onset and monitored morning levels of salivary cortisol. Ortiz-Tudela et al. (2010) produce a composite metric of WT, activity and body position which they suggest can characterise circadian rhythms. The authors suggest that this metric can reveal the links between circadian disruption and conditions such as metabolic syndrome, diabetes, cardiovascular disease and even cancer prognosis. However, this metric is complex and not readily reproducible in real world settings.

Romeijn and va Someren (2011) found WT was inversely proportional to vigilance, as measured through a clicking response task. This finding correlates with other observations that WT increases with sleep periods. Ortiz-Tudela et al. (2014) use their metric to show that phase advances occur in the biological clock of patients with Mild Cognitive Impairment. Using data from 46 healthy individuals in the Netherlands, Vinkers et al. (2013) found that stress did not change wrist temperature significantly, even though it does effect the temperature readings at other regions of the body. This is a useful observation, and it provides further evidence that a wrist worn temperature sensor may be the most practical position for a temperature sensor worn on the body to best characterise the immediate thermal environment of the wearer.

This final section of the literature review has revealed that wrist worn temperature monitors have not yet been deployed as a means of characterising the experienced temperature of the wearer. Moreover, the comprehensive review of wrist temperature variations reveals that there is no evidence of substantial effects which would significantly impact a wrist worn monitor as a function of demographics. The clearest effects are the diurnal variation in wrist temperature, which suggest wrist temperature is highest at night-time. There is evidence to suggest that obesity and metabolic syndrome reduces the amplitude of wrist temperature variation. There is also clear evidence that disrupted circadian rhythms are detectable through wrist temperature measurements. However, at a population level it would possible to control for both obesity levels and shift-workers.

2.6 Summary

This review sought to examine the literature at three different levels, all of which are relevant to this study's research aims. At the largest scale, clear evidence for the winter peal in mortality is evident. This is most likely related to cold external temperatures, although the impact of covariates such as air pollution is also important. It is unclear from the epidemiological evidence at this level as to where, when and for whom this harmful exposure occurs. As people spend over half their lives at home, it is likely that household temperatures play a significant role in this regard. Exposure to cold in the work place is less likely, in large part due to the standardisation of indoor work environments, which according to the Health and Safety Executive should be at least 16°C (HSE, 2007).

At the level of the household, differences in internal temperatures are observed. There is evidence to suggest that Local Authority housing is warmer than privately owned homes (Hamilton et al., 2017; Kelly et al., 2013). Older people and people with long-term disabilities are found to have warmer dwelling temperatures than younger people and those without long-term disabilities (Huebner et al., 2018). However, there is concern in the literature that occupancy may not line up with the measured differences in temperature – it is therefore useful to consider monitoring the immediate thermal environment of different demographic groups to understand how dangerous cold exposure occurs.

There has been little attention given to the problem of measuring person cold exposure at the population level. The systematic review of wrist temperature variation showed little variation in wrist temperature as a function of demographics. Using a wrist worn temperature monitor is a feasible method of monitoring the experienced temperature of individuals (this is explored in detail in chapter 5). The following chapter discusses the UK Biobank in detail and synthesises the evidence of this review to develop the study research questions and hypotheses.

Chapter 3

UK Biobank and the conceptual model of the study

My health is better: I lay the blame of its feebleness on the cold weather, more than on an uneasy mind.

Elizabeth Gaskell – The Life of Charlotte Brontë (1857)

The previous chapter identified a significant gap in the literature regarding the understanding of the relationship between the temperature directly experienced by an individual during their everyday life and their health outcomes. In order to explore this theme, this study makes use of data from the UK Biobank as well as a wrist worn activity and temperature monitor, the AX3. This chapter describes the UK Biobank more fully, and outlines the conceptual model which underlies this thesis.

3.1 The UK Biobank

As described in the introductory chapter, the UK Biobank is an on-going prospective cohort health study of UK adults, which recruited over 500,000 people from the general population. The cohort was aged between 40 and 69 years when recruited between 2006 and 2010 (Sudlow et al., 2015). The diversity of data available for researchers is striking and includes both genotypic and phenotypic information collected using a wide range of methods. The data are available to any researchers following a detailed application process. The application for this study was approved in May 2017. Following this, a Material Transfer Agreement was signed, committing the project to strict data storage protocols (see section 4.1 for more details). The overall data collection time frame of the UK Biobank is given in table 3.1. This study is concerned primarily with the data collected by the AX3, but the relationship between these data and the sociodemographic data collected throughout the UK Biobank study is also very important.

Of specific interest to this study were the 236,519 participants who were invited to wear

Name	Description	Start	End
Pilot phase	Tested the entire recruitment	03/2006	06/2006
Baseline assessment	Sociodemographic, lifestyle, health, environmental, and psychosocial factors were collected using questionnaires, tests and biological samples	04/2007	10/2010
Diet	Collected using a dietary recall questionnaire	02/2011	06/2012
Baseline assessment repetition	A repeat of the above baseline data collection	08/2012	06/2013
Genomics	Genome-wide genotyping from blood sampling on all participants	2013	2015
Registries	Participants are linked to death and cancer registries and inpatient records.	03/2013	2017
AX3 physical activity monitor	The primary data for this study, described in the text, collected for 100,000 participants	05/2013	12/2015
Cognitive function	A web questionnaire collected for 100,000 participants	11/2014	12/2014
Occupational history	A web questionnaire collected for 100,000 participants	07/2015	09/2015
Imaging	Multimodal body imaging data collected for 100,000 participants	10/2015	2017
Mental health	A web questionnaire collected for 100,000 participants	08/2016	12/2016

Table 3.1: The data collection time scale for UK Biobank. Adapted from information summarised by Maelstrom (2019)

an AX3 physical activity wristband sensor for one week, between May 2013 and December 2015. The AX3 encodes accelerometer, temperature and light level data into a single data file. 103,707 such files were available for analysis in this study. The literature review identified the need for a study examining the experienced temperature of individuals in relation to health and energy demand. The following section assesses the suitability of using the AX3 wristband as a means to measure the experienced temperature. Three principle experiments are described. Following this, the conceptual model for experienced temperature which underpins this study is given.

3.2 The Axivity wristband sensor

The Axivity AX3 wristband was designed to measure participant activity levels using an accelerometer which samples acceleration at 100Hz. The accelerometer's performance varies with temperature, and so must be normalised, although these effects only become important at high temperature (Axivity, 2015). In order to monitor the temperature of the device, a MCP9700 thermistor is included on board the Axivity device (Microchip Technology, 2007). It samples temperature with a period between 1.1 to 1.3 seconds. This non-constant data

Figure 3.1: The Axivity AX3 wristband, worn while painting.
Image source: `https://bit.ly/2LsXqTf`

collection period results from the way the device collects data. It was therefore necessary to process and downsample the collected data, in the first instance to create a uniform sampling time, and in the second to make data more manageable. This served to reduce the file size for each participant and made producing the regression models more straightforward. This process is described in more detail in section 4.1.

The activity information which the AX3 records was processed to a 5-second period, from an initial rate of 100Hz. Very little meaningful information about a participant's experienced temperature exists below the 5-second period, and indeed home monitoring programs tend to operate at a minimum period of 5 minutes (see section 4.2.2). Therefore, a period of 1-minute was chosen for the final downsampled data, for both the temperature and activity data. This is a conservative compromise between file size and potential information loss. The file format for data from the Axivity wristband is CWA. For a week of data, the CWA files are around 260 MB each. The whole unprocessed data set for all 103,707 participants is therefore 27 TB. Processed to a 5-second period each file is approximately 13 MB, which makes the whole data set 1.3 TB. The further downsampling to a 1-minute period made the entire dataset 165 GB in all.

Parameter	Measurement
Dimensions	23 x 32.5 x 7.6 (mm)
Weight	11g
Moisture Ingress	IPx8 1.5m for 1hr
Dust Ingress	IP6x
Memory	512Mb flash non-volatile
Accelerometer	Sample Rate 12.5 - 3200Hz Configurable
Battery Life	30 days @ 12.5Hz; 14 days @ 100Hz
Accelerometer Range	2 / 4 / 8 / 16g Configurable
Accelerometer Resolution	Up to 13 bit
Thermistor frequency	1.1-1.3s
Thermistor response time	Similar to HOBO monitor
Thermistor accuracy	$\pm1°$under standard operating conditions
Thermistor resolution	0.3°

Table 3.2: The physical parameters and functionality of the AX3 (Axivity, 2015; Microchip Technology, 2007)

It is important to note that the AX3 was not designed as a temperature monitor per se. The accelerometer's functionality can be impacted by the device temperature, although usually not under normal operating conditions. The on-board thermistor's primary function is to provide calibration for the accelerometer. However, as will be shown in chapter 5, the AX3 can be used successfully to measure temperature which reflects the ambient environment and the heat from the wearer's wrist.

3.3 The conceptual model of experienced temperature

In general, the thermal environment of humans is highly complex (Gagge and Nishi, 2011). The primary pathways for heat transfer are conduction, convection, radiation and evaporative exchange. Since these pathways can be highly anisotropic and heavily influenced by local heat flows, clothing, air movement, moisture levels and many other factors, characterising this environment is difficult. No single temperature will capture an individual's thermal environment fully.

However, a wrist worn temperature sensor reflects the thermal environment well because it is generally further away from the core. As depicted in figure 3.2, when an individual is in a cold environment, blood flow is restricted from the extremities, which means a wrist worn sensor reads a colder temperature in line with the ambient environmental temperature.

Wrist temperature varies diurnally with a peak during sleeping hours. However, as was shown in chapter 2.5, there is little variation in amplitude reported as a function of either gender or age. Therefore, recorded differences between demographic groups likely correspond to environmental differences between these groups, all other things being equal.

The complicating factor of clothing should also be considered. As demonstrated in figure 3.3, the addition of layers of clothing worn over the AX3 increases the temperature reading relative to an uncovered AX3. An ambiguity therefore exists as to whether a warm

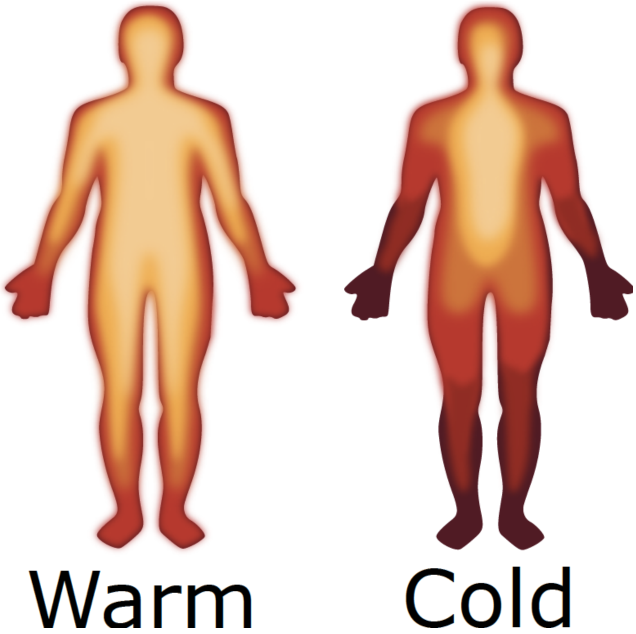


Figure 3.2: The approximate distribution of body heat in warm and cold conditions. Generally, in colder conditions, blood flow to the extremities is restricted to reduce heat loss. Dark areas indicate temperatures around 30°C, light areas indicate 37°C. The exact meaning of warm and cold conditions depends on clothing levels, metabolism, gender and many other factors. Adapted from White et al. (2011)

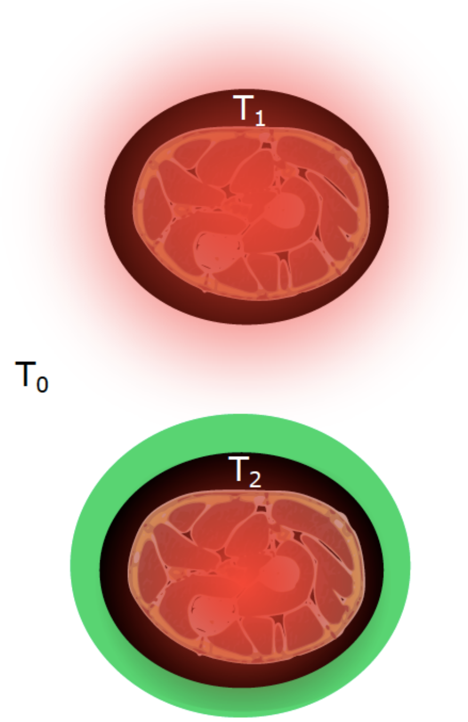


Figure 3.3: The AX3 registers a warmer reading when worn under clothing than on top of it, even in the same external environment

reading (e.g. above 30°C) is the result of a warm room, or a sensor worn under clothing. The same is not true for the converse situation: a low reading (e.g. below 20 °C) corresponds to a cold environment, as there are no physical mechanisms which allow the wrist to be colder than the environment in cold conditions. It is therefore not possible for a cold reading on the AX3 to be the result of anything other than the participant being in a cold environment. This minimises the risk of a Type I error regarding whether the participant was exposed to cold. Type II errors are more difficult to discount, since it is plausible that a participant might have breathed cold air while wearing an AX3 that was covered by clothing. However, it is probable that such instances were rare given the 5-day monitoring period.

Following on from the gaps identified in the literature, and the conceptual model outlined above, the following section sets out both the preliminary and main research questions for this study.

3.4 Research questions and hypotheses

Two preliminary research questions, denoted P1 and P2, seek to answer the following questions:

P1. Does the AX3 function accurately as a temperature recording device?

P2. If so, to what extent does it record the immediate thermal environment of the wearer while being worn?

The AX3 device pilot testing, which is described in full in chapter 5, shows that the AX3 records temperature with sufficient accuracy to allow for the characterisation of experienced temperature. The specific tests that were carried out, their rationale, and the hypotheses which were formulated prior to them being carried out are also addressed in chapter 5. The main research questions that are addressed by this thesis are more general than P1 and P2, but answering them in the manner described in this thesis logically requires that the AX3 is shown to be an adequate temperature recording device. Therefore, the results to P1 and P2 are given prior to the main results section. The main research questions of this thesis are as follows:

RQ1. Does experienced temperature vary with sociodemographic and building variables [e.g. sex, age, ethnicity, income, building type, tenure]?

RQ2. Are there associations between experienced temperature and the health conditions related to excess winter deaths, i.e. cardiovascular and respiratory diseases [ICD 10 I00 – I99 and J00-J99]?

RQ3. Do combinations of sociodemographic factors, building factors and the health conditions related to excess winter deaths (as above) have associations with low experienced temperature?

Following on from these research questions, the subsequent hypotheses are given, based on the evidence available in the literature. These hypotheses will be revisited in chapter 9 and critically discussed based on the findings reported in chapters 6, 7 and 8. The hypotheses below relate to the dataset for which shift–workers, those with health conditions that lead to unusually cold hands, and those with dementia and Alzheimer's disease have been excluded. This is discussed further in section 4.8.

Hyp1. No significant differences in experienced temperature will be measured as a function of sex, ethnicity and income. Experienced temperature will increase as a function of increasing age and decreasing health satisfaction. Experienced temperature will be higher for those who live in Local Authority housing than those who live in owner-occupied accommodation.

Hyp2. Those who have health conditions associated with excess winter deaths will be more likely to have higher experienced temperature.

Hyp3. Those who have health conditions associated with excess winter deaths will not be more likely to have higher experienced temperature if they are also in low income households.

3.4.1 Discussion

Hypothesis 1 follows from the reading of the literature that there are no significant differences in wrist temperature as a function of demographics. Therefore, the observed differences will be due to ambient conditions. The literature review showed the clearest evidence for increased ambient temperatures as a function of increasing age, and for those who live in Local Authority housing. There was also evidence which showed that those with long term disabilities had higher domestic temperatures (details of exactly how the variables are operationalised is given in chapter 4).

Hypothesis 2 again follows from the evidence that those with long term disabilities have higher domestic temperatures. These participants likely increase their domestic temperature, and consequently their experienced temperature, in order to alleviate the symptoms of chronic health conditions (such as respiratory problems). However, others may not have such the financial capability to do so. Therefore, hypothesis 3 follows from the assumption that participants who are less able to afford home heating, i.e. those whose income is low, might ration home heating. This final hypothesis is made with less direct evidence from the literature, although there is evidence to suggest that rationing of heating does occur in certain low–income households.

3.5 Pre-analysis plan

As was shown above, the number of available variables in the UK Biobank is very high. *Data dredging* or *p-hacking* is the practice of searching large databases to find statistically significant relationships which are identified only because of the databases size, and not because of some real underlying relationship. In order to guard against this, the variables of interest and statistical tests undertaken in this thesis were pre-specified in a Pre–analysis plan (PAP). This was uploaded to an online repository prior to the data analysis stage (Kennard et al., 2017) and is reproduced in appendix E. However, slight modification of the research design was required following additional literature reviews. The language of the research questions was also improved relative to what was published, since statements in the original version lacked clarity.

The next chapter goes into detail about the specific ways in which the main research questions given above will be addressed in the study. The results of the main research questions are given in chapters 6, 7 and 8.

Chapter 4

Method

. . . it is extremely desirable that Bed and Sitting Rooms for Winter occupation, should have a Southern aspect – when the Thermometer is below 30, the proper place for people beyond 60, is their own Fire-side. . .

William Kitchiner – The Art of Invigorating and Prolonging Life (1822)

4.1 Methodology

The first section of this chapter discusses the methodology of the study as the whole. Here, methodology is taken to mean a collection of methods deployed to address a research question (Kothari, 2012). The main research questions outlined in the previous chapter lend themselves most appropriately to a quantitative methodology. This choice is partly based on the constraints of the broader research design of the UK Biobank, which was set up with a purely quantitative structure in mind. However, it is also the case that qualitative data is very challenging to collect and synthesise at the scale of 100,000 participants. Moreover, the primary variable of interest is temperature, which makes quantitative methods appropriate.

The nature of the PhD limits the possibilities for measuring experienced temperature at a population level – it would have been impossible to deploy thousands of sensors within the timescale and budget of the study. The previous chapter laid out the conceptual model of why a wrist worn sensor is a good method of measuring experienced temperature. Alternatives might have been an ankle work sensor, or an iButton worn on clothing (cf. the method of Kuras et al. (2015)). Following this, the most appropriate methodology to analyse the data collected by the AX3 is that of an epidemiological approach. For this, multiple-regression models estimate the strength of associations between variables. As mentioned in the previous chapter, the core of the study was pre–specified in the Pre–analysis plan (PAP), which is reproduced in appendix E and was published online before the data were analysed (Kennard et al., 2017). Following this, extensive pilot

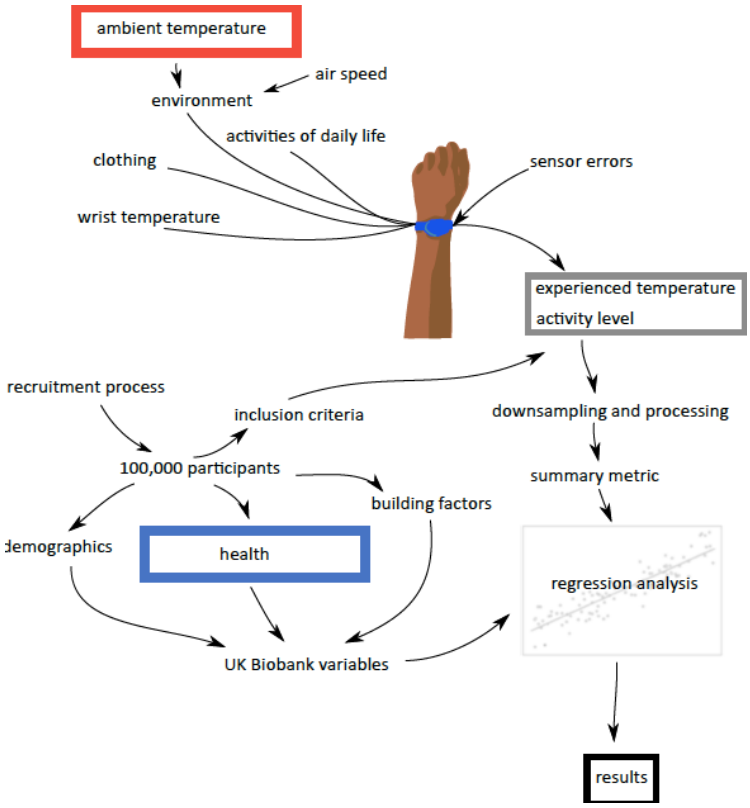


Figure 4.1: An overview of the study as a whole. Ambient temperature and health are highlighted as they are of primary interest. Other factors and processes which influence the results are shown

testing was conducted in order to answer the preliminary research questions given in the previous chapter pertaining to whether the AX3 records temperature accurately and whether the temperature it records reflects the immediate thermal environment of the wearer. These tests are discussed in full, including the methods used, in the following chapter. The next chapter also goes into more detail about how the AX3 compares to using an iButton. The conclusion from these tests was that the AX3 is sufficiently accurate at measuring temperature, and that the temperature it measures reflects the thermal environment of the wearer. The remainder of this chapter is therefore devoted to describing the methods for the main analysis of this thesis. As suggested above, the most appropriate method for determining the relationship between variables in epidemiology is using regression. However, before discussing the regression models, the variables must be introduced, and most importantly, the method by which the experienced temperature variable is derived. This portion of the study involved processing over 27 TB of data. Designing and implementing the data processing workflow took just under half of the available time for the research portion of the PhD, which was around 30 months in total.

The methods of collecting, processing and analysing the information available to this study are represented in figure 4.1. Here, the term information is used in an abstract sense – it refers to the variety of ways data were collected and analysed. Two key input sources of information are highlighted in figure 4.1, namely the ambient temperature of a participant and their general health status. The diagram aims to capture how these variables were operationalised, both through the concept of experienced temperature and the UK Biobank variables in general, before being passed to the regression analysis stage, which ultimately led to the results of the thesis. Separating the information flow in this way helps to summarise what has been carried out and provides a structure to aid the explanation.

The following section describes how the UK Biobank data were processed. However, prior to this, access to the data resource had to be gained. The study began with the supervisory team identifying that the UK Biobank resource and the temperature variable recorded by the AX3 would constitute an interesting program of study. The data were initially identified through the UK Biobank Data Showcase (UK Biobank, 2019). The variable of primary interest at this early stage was the average temperature recorded by the AX3 (which was erroneously given the value 20.9°C by the UK Biobank team, as is discussed in detail in section 5.1 of the next chapter.)

The work began with a formal application for data access to the UK Biobank, which was done in conjunction with the supervisory team. This required setting out the research plan in detail, and providing a list of all required variables. This list was longer than those which were eventually included in the PAP (see appendix E), because the final list of variables had not been determined at that early stage.

4.2 CWA Processing

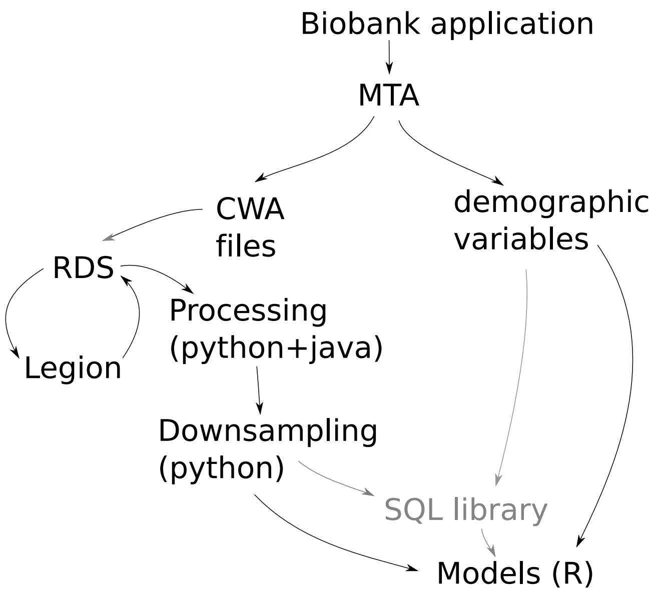


Figure 4.2: The CWA processing pathway. The processing and downsampling stages were the most time intensive aspects of the study. The Material Transfer Agreement (MTA) and Research Data Services (RDS) storage space are also indicated.

Once the application was submitted, work on determining the minimal set of variables which would allow the research questions to be answered began. This required consultation of the literature, and clear understanding of the nature of the Biobank variables. One example of a variable that was initially requested, but then not made use of, was the Townsend Deprivation index. This was because the index was only defined at Census Output Area level, which have a minimum of 40 households or 100 residents (ONS, 2019), and was therefore not specific enough to the participants of this study to be of use. Other unused variables were included in order to facilitate potential further research which fell outside the scope of this thesis. As will be described further in section 4.6, further literature review work and consultation with colleagues who had more medical expertise revealed that the inclusion of three further variables would be required.

The application process proceeded in two stages. It was approved in principle following an initial application in January 17$^{\text{th}}$ 2017, and in full on May 8$^{\text{th}}$ 2017. Following this, a formal Material Transfer Agreement (MTA) was signed on September 18$^{\text{th}}$ 2017 (UK Biobank, 2017). This required accepting the condition that the data would be held in a secure system. The system where the majority of the primary data were stored was password protected with encryption, with the password known only to the author. The UK Biobank data is anonymised, so determining the identity of the participants is generally not possible. Exceptions to this were considered, for example, instances in which a person with specific sociodemographic characteristics living in an area of low population. However, since the home location variable is rounded, it would be extremely challenging to identify such an individual, it was deemed to be an acceptably low risk. At this stage an ethics audit was conducted which required the study to be outlined to the UCL Energy Department lead on Ethics. Since the UK Biobank data are anonymised, it was concluded that a separate application to the UCL Ethics board was not required for the study.

The data were divided into two portions. The first comprised the sociodemographic, health and housing data of the participants. These were around half a GB in size, so were able to be stored on the local secure system. The second portion comprised the CWA files were much larger. Each CWA file was around 250 MB, which meant the whole dataset was over 27 TB in size. This required the use of the UCL Research Data Systems to store them securely. At this stage, Dr Stuart Grieve and Raquel Alegre were taken on using funding from the Centre of Energy Epidemiology (CEE) budget. They provided one week's worth of support for the specific computing tasks required to process the CWA files, and they also assisted in refactoring the Python and Java scripts that were written by the Axivity team at Newcastle University (NU) to process the CWA files. The modified scripts are available at `https://github.com/UCL/AX3-temp-output`. Furthermore, they assisted with teaching the necessary SQL programming techniques to build the database which held both the sociodemographic data and the processed CWA data. However, as is indicated in figure 4.2, this step was later deemed to be superfluous as the R programming language was sufficient to allow direct incorporation of the sociodemographic data and the summary metric of the processed CWA files. It was initially expected that this would not be possible, but as the research design evolved it became clear that the complete database of sociodemographic and summary metric data would be small enough for R to be able to process directly.

4.2.1 Processing CWA files

Processing of the CWA files was divided into two distinct phases. Step one was the initial processing which decodes the CWA and converts it into a CSV file. This was conducted using the UCL computing cluster with the modified NU Python and Java scripts mentioned above. The second stage, once the CSV file was produced, involved further downsampling to a 1-minute period. This work was conducted on the local system using Python scripts (see below section 4.2.2). These time-series were then summarised into a number of different metrics, as described below in section 4.3.1.

The details of the workflow for step one using the UCL computing clusters are given in appendix A.1 which was created by Stuart Grieve. The workflow is an idealised representation of the practical process, which included multiple technical drawbacks and complications. These difficulties are described in the next section. The workflow was created with assistance from Stuart Grieve and Raquel Alegre. The broad scheme was as follows. First the unprocessed CWA files were downloaded using the UK Biobank data downloading tool, which was automated using a BASH script. Then, the processing script was compiled in a virtual environment on the computing cluster. The virtual environment acts like an isolated operating system that has very specific software libraries that hold the code that the script requires. This step was complex as the NU script required specific libraries which were not available by default on the UCL cluster. Once the virtual environment was established and the NU script was running correctly, the CWA were copied securely (using the SCP command) from the RDS storage to the cluster. These files were then processed in batches using the NU script, up to a maximum of 100 at any one time (depending on how many other users were running jobs). Each CWA file took around 1 minute 40 seconds to process. It was this batch processing that allowed the totality of the 100,000 participant files to be processed within the time frame of the PhD. Without batch processing the CWA files would have taken over 3 months to process. Once each file was processed the resultant CSV files and JSON files, which contained the default summaries of wear time and average temperature as well as notification of any errors, were sent by SCP back to the RDS storage system. The NU script developed by Doherty et al. (2017) was designed to calculate average activity for every hour using the AX3. The AX3 is a triaxial activity monitor, which means it is sensitive to acceleration in all three orthogonal directions. It is configurable in a number of ways; Doherty et al. (2017) chose to sample acceleration over a seven day period at 100Hz - the range of possible acceleration measured by the device were $\pm 8g$ (where g is the acceleration due to gravity). The device is capable of sampling at higher rates, which allow high intensity activities, such as boxing, to be captured, but this compromises battery life. Stationary episode were defined as periods when the accelerometer measured a standard deviation of less than 13.0 mg (Doherty et al., 2017).

4.2.1.1 Difficulties

Processing 27 TB of data is challenging, even using modern computing methods. Most of the difficulties encountered arose from simple errors in the processing scripts which were rectified by carefully examining the code. For example, the file path of a process not being properly defined would prevent a process completing. The nature of the cluster system meant such errors were not always easy to track down. Other setbacks occurred due to system shutdowns or overuse and so were outside the control of the author.

A more substantive error in the NU script was discovered several months after the entire process had been completed. The error affected around 5% of the processed files, and so was not evident at first. It was caused by a single number being incorrect in one line of NU script (which is over 900 lines long). This error was the result of the refactoring of the script during the early stages of the processing exercise. The number in question referred to the column number of the CWA data – for the 5% of files affected the activity and temperature columns would be exchanged and the temperature readings replaced by nonsensical activity readings. The discovery and correction of this error took around a week. Unfortunately, there was no way of determining which files had been affected, since the error did not appear on any of the output error files, and the entire dataset had to be re-processed. UCL Research Data Services (RDS) had brought on-line a new cluster system named Myriad, which was in the early testing stages. This meant no other researchers were using it at the time and the whole data set was reprocessed in 3 days - this was far quicker than the original processing time.

4.2.2 Downsampling

The resultant time-series CSV files containing temperature, activity and light level readings were sampled at a rate of once every 5-seconds. As will be discussed in chapter 5 on the pilot studies, the light level readings were too inaccurate to be of use. The 5-second sampling period was deemed unnecessarily short for practical use in the research project. Most physiologically relevant processes to this study take place on a time scale longer than this 5-second period. 1-minute was chosen as it created a more manageable file size, but still contained a sufficiently fine temporal resolution – domestic temperature studies use a variety of sampling periods, typically between 5 and 45 minutes (e.g. Kane (2013) uses 5 minutes; French et al. (2007) use 10 minutes; Hamilton et al. (2017) use 20 minutes; K Firth and Wright (2008) use 45 minutes) . The downsampling to 1-minute period was achieved using a Python script (given in appendix C.1). This stage also converted the wear time summary that is given as one of the outputs of the NU script to a simple binary variable (0 for unworn and 1 for worn) for each time-stamp. This meant all the relevant information was contained in a single file, which made the summary stage described in the next section much more straight forward.

Downsampling always involves removing information from a time-series. If conducted properly, the information removed is of little or no value, and serves to remove random fluctuations from the data resulting in time-series of lower temporal resolution that has been smoothed out. Correct downsampling procedure ensures that measures such as the mean value of the data are not affected by the smoothing. Tests were conducted during this process to ensure the Python algorithm was accurate in these regards. This was done in two ways. Firstly, a simple method of plotting both the downsampled and original time-series on the same axes to see that they followed the same pattern. This was followed up by calculating the overall average temperature reading for a particular test example – all tested files had near identical averages, agreeing to at least 2 d.p. Downsampling reduces the level of noise in a time-series, and so very high frequency variations are averaged out. This can result in a mild reduction of range in the downsampled time-series. It was ensured that the downsampling algorithm did not result in dramatic range reductions. In the spot checks conducted, no range was reduced by more than 1°C, and most were reduced by around 0.1°C.

4.3 Experienced temperature

The final data comprised a 1-minute period time-series of temperature readings from the AX3 as well as the average activity level (acceleration measured in units of g) recorded for that minute. For each minute in the time-series an estimation of whether the AX3 was being worn was also given, as a binary variable. The combination of these data allowed the exposure metric to be derived, which is described in the following section.

4.3.1 Exposure metrics

As mentioned above, once the 1-minute period time-series of temperature readings were produced, they were summarised into the various metrics that this study uses. The script which produced these summaries is given in appendix C.2. The most straightforward way to input the experienced temperature into a regression model is to summarise it, that is, convert it from a time-series of 7,200 minutes into a single value which captures the characteristics of the data (the 7,200 minute time-series results from the five included days, measured at minutely intervals, see section 4.3.1.1). This can be done in a number of different ways. The most intuitive is the mean. Examples of other summary metrics of time-series are the median, maximum or minimum, or the standard deviation. The details of the metrics are given in table 4.1, which includes a subscript notation used throughout the remainder of the thesis. There are several reasons why such alternative summary metrics should be taken into account. First, since most people spend around 8 hours per day in the warm micro-climate of the bed, the mean is likely dominated by this temperature, which is relatively static. Second, as pointed out in the conceptual model in chapter 3, using an AX3 to measure cold exposure is ambiguous at warmer temperatures, due to the argument that a warm reading on the AX3 could either be because of a warm environment or layers of clothing over the device in a colder environment.

Metric	Description	Notation
minimum	the lowest value in the data	t_{min}
first decile	the value corresponding to the lowest 10% of the data	t_{10}
lower quartile	the value corresponding to the lowest 25% of the data	t_{25}
mean	the first moment of central tendency	t_{μ}
median	the value separating the upper and lower half of the data	t_{50}
upper quartile	the value corresponding to the lowest 75% of the data	t_{75}
ninth decile	the value corresponding to the lowest 90% of the data	t_{90}
maximum	the highest value in the data	t_{max}
standard deviation	a measure of dispersion	t_{sd}

Table 4.1: A table of some of the some of the potential summary metrics of a time-series Bowley (1920)

The argument was made in the PAP that the first decile be chosen as a summary metric. However, the decision was made that a single metric may be too restrictive in the final analysis. Therefore, the potential metrics (given in table 4.1) were divided into two types. These were the *lower experienced temperature* metrics and measures of *thermal variety*. The lower experienced temperature metrics are those which aim to characterise cold exposure, as described in the conceptual model. Specifically, these are the minimum recorded temperature and the first decile. However, the minimum is potentially impacted by outliers as it is a reading which only corresponds to a single minute of exposure. By contrast, the first decile corresponds to a reading of 12 hour's worth of exposure (i.e. 10% of the total 7,200 minute monitoring period). The two metrics are compared in the results chapter which follows. There is no a priori justification for taking the first decile over, for example, over the 5$^{\text{th}}$ or the 15$^{\text{th}}$ percentile, other than it providing a good balance between capturing periods spent in cold conditions, which might not be picked up by the average, and the minimum temperature, which might be the result of a single brief minute–long cold exposure that is not representative of the participants experienced temperature as a whole. It was expected that between these two variables the cold exposure would be captured (subject to the caveat regarding clothing outlined in the previous chapter). It is important to stress that these arguments were made on the basis of the pilot studies alone, prior to analysis of the general UK Biobank data set. The mean itself was included as a metric since it is a very widely used summary metric. Even though it likely does not characterise cold exposure per se, it could be argued that having a lower average temperature could be the mode of action for the harmful effects of cold to be felt. These considerations will be discussed further in the discussion chapter 9.

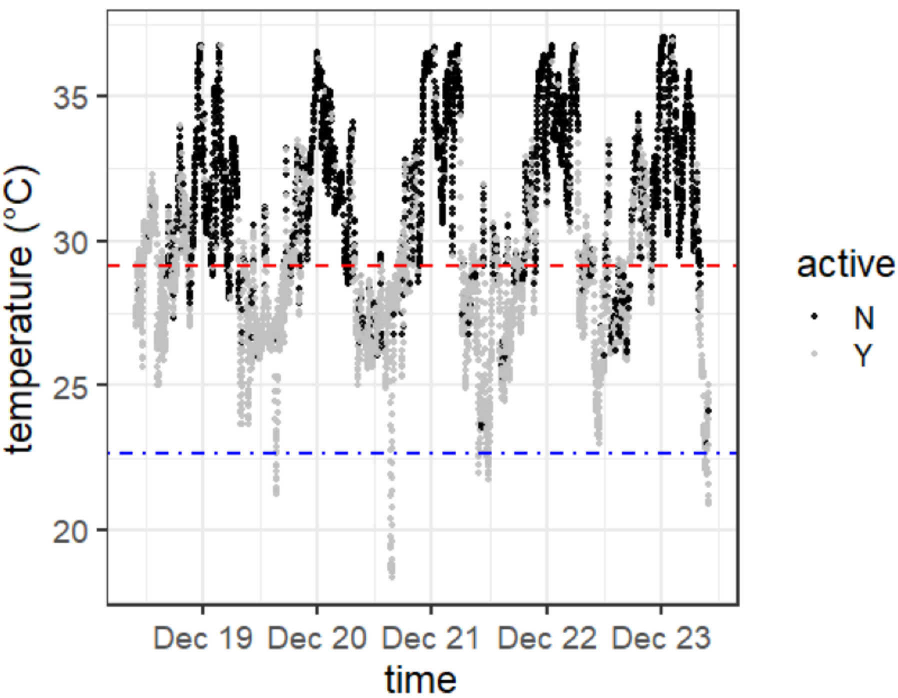
The second kind of metric is the thermal variety, defined in this study as the standard deviation of experienced temperature. This aims to capture the diversity of the temperatures that the participant was exposed to. Unlike the first decile it uses the whole time-series in its calculation. The expectation was that the greatest diversity of readings would result from different levels of cold exposure. That is to say, the bedroom and sleep related readings were expected to be relatively similar between participants. For this reason, the thermal variety was thought to be dominated by cold exposure, and therefore worth investigating. The outline of the following chapters matches this division of metric types: chapter 6 focuses on the lower metrics, and 7 on the thermal variety. The metrics included are given in table 4.2. The next consideration for the metrics, discussed below, regards the inclusion of the activity data.

Metric	Notation	Comments
Minimum	t_{min}	the coldest value recorded during the study period
First decile	t_{10}	the value corresponding to the lowest 10% of the data
Mean	t_{μ}	The mean was not expected to be a good means of characterising cold exposure, but is included for completeness
Standard deviation	t_{sd}	a common measure of dispersion, expected to characterise cold exposure

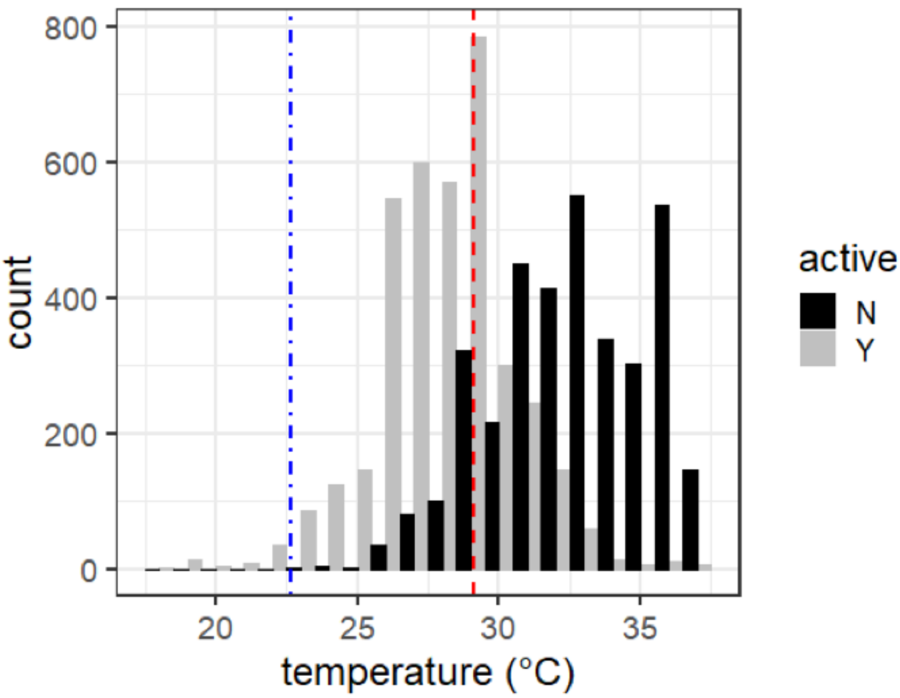
Table 4.2: The summary metrics used in the thesis, the notation used and a description of what they measure.

4.3.1.1 Activity

A second consideration for the data summary method was how to incorporate activity data into the analysis. The central problem arises because there is no information available which indicates when a participant is indoors. This means it has to be estimated in a different way. For this, all periods of time where participant's activity was above their median activity level was excluded. In order to denote this, a superscript 'm' is added to the notation given in table 4.1. For example, the first decile screened by median activity level is denoted t_{10}^{m}. When no exclusions to the time-series are conducted, no superscript is given. An example of a temperature time-series, summarised by the minimum and first decile is given in figure 4.3a, and the same data represented as a histogram in figure 4.3b.



(a) Example participant experienced temperature time-series, shaded by whether they were active or not (Y or N). Higher night-time measurements correspond to both the warm micro-climate of the bed and the mildly increased temperature of the wrist during sleep.



(b) Experienced temperature histograms shaded by whether the participant was active or not (Y or N).

Figure 4.3: Example experienced temperature time-series and histogram. In both figures, the dotdash line denotes t^m_{min} of the data, and the dashed line denotes t^m_{10} (the first decile).

This decision was made for two principle reasons. Under the assumption that little time is spent outside while sedentary during cold periods of the year, the lower activity periods are more likely to reflect internal environments. Since this study is primarily motivated by energy use in buildings, it was decided that focusing on periods for which the participants were likely inside was most appropriate. Second, since the metabolic response of people is higher during periods of moderate activity, the negative impacts of cold are likely to be greater during sedentary periods. Both of these assumptions will be critiqued in section 6.4.

For the study due to Doherty et al. (2017), which produced the raw data processed and analysed in this study, participants were asked to wear the device continuously, and put it on as soon as they received it in the mail. It was programmed to activate two days after the postal date. Participants were asked to return the device at the end of the 7-day wear period using a pre-paid envelope. The devices were calibrated using industry standard procedures to local gravity levels. Periods where the device was not moving (accelerating) were identified by a 10-second window during which all three axes of the device read less than 13mg. These stationary periods were then used to optimise the local gravity readings which were subtracted from the final acceleration readings. For instances when the device was never determined to be stationary, this calibration was determined by the next time the device was used (since the same device was used to measure multiple participants activity consecutively). The acceleration signal was then normalised, by computing the Euclidean norm of each axis reading, finally a fourth order Butterworth low pass filter was used to remove noise. These individual sampling periods were then combined into five second *epochs* – these determined the output period of the processing described above. A non-wear period was then defined as consecutive stationary periods lasting for 60 minutes. It is for this reason that temporary removal of the device for a very short period would not be registered as stationary, which is discussed below. Activity levels during non-wear periods were imputed using the average of the similar time of day from other measurement periods. Since the present study uses fundamentally the same script to analyse the CWA data, the processed data are equivalent, with exceptions that will be noted. During the investigation to determine whether the AX3 accurately records temperature, described in the following chapter, it became clear that that the algorithm that determines whether the device is worn or not are not 100% accurate. This is because it relies only on detecting device motion, so in effect a wear-period is detected irrespective of whether it actually being worn, but rather whether the device is moving. In terms of this study, issues occurred particularly at the start and end periods of the wear time. Since the device was programmed to automatically start at the beginning of the monitoring period, it activated whether or not it was actually being worn. Although no direct evidence of this was available, it is likely that some participants experienced postal delays that prevented them from wearing the device. This would have meant the device could have been switched on and recording temperature and activity data. If during the postal process it was moved this would lead the device to assess such times intervals as being worn, when in fact they were in transit. The evidence for this was that the AX3 temperature reading would be around 20°C whilst also being registered as worn - given that heat from the wrist is also detected from the device, readings of 20°C are rare, especially given the relatively mild climate conditions of the UK. Therefore, as a conservative precaution, the first and last days of each participant were removed, even if the requirement of a 90% wear-time was satisfied. This reduced the likelihood of such errors effecting the study.

4.4 Measurement

It is helpful to review exactly what is measured by the AX3 device, and how well this approximates the variables of interest. This section corresponds to the top portion of figure 4.1. The AX3 device records temperature with an accuracy of approximately ±1°C. One key assumption of the analysis is that these errors are distributed at random across the AX3 devices used in the study. Once participants are grouped by sociodemographic data, for example, these errors cancel each other out, and the estimates produced by regression analysis are centred around the true values. This is a reasonable assumption given that the AX3 devices were randomly distributed to participants.

A second type of sensor error was encountered during processing of the data. When examining the time-series of the outliers of t_{10}^m in detail, it was found that those with very low or very high readings (i.e. < 20°C or > 40°C) had several issues. One such issue was that of the catastrophic sensor failure, for which the temperature sensor produced readings that varied so widely that they could not conceivably correspond to any physical circumstances. A second failure were brief periods where the temperature times series dropped sharply and then recovered to the previous value – these periods likely corresponded to time periods when the sensor was removed, but the accelerometer processing algorithm did not detect them as such. The method for determining these exclusions was as follows. First, using the t_{10}^m metric, the time-series of outliers (i.e. < 20°C or > 40°C) were saved to a file. Each of these files was then inspected to understand why the particular time-series had extremal values. The reason was noted in a separate log file which gave the participant ID number and the particular kind of error (i.e. apparent sensor failure or brief flat lining of the temperature time-series). The resultant participant ID numbers were then excluded from the dataset. This process was repeated for t_{min} and t_{sd}, with a total of 118 such exclusions made. The overall impact of these exclusions was very small, and did not impact the findings of the study. This is most likely because the outliers were randomly distributed within the data.

Given the evidence that the AX3 faithfully records the device temperature, which is described fully in the following chapter, it is useful to consider the impact of temperature fluctuations in the region of the wrist which are not correlated to the overall thermal environment of the wearer. Examples of such sources of local atypical temperature fluctuations are clothing, circadian rhythms, incident radiation, local air movement and water immersion. Without primary data on clothing level on the wrist, such effects are all but impossible to quantify or account for. In the next section the temperature time-series recorded by the AX3 are considered, with a view to determining whether any evidence of these effects exists in the data.

4.4.1 Temperature time-series

Since there were 101,801 time–series which were successfully downsampled to 1–minute period, it would not have been possible to inspect them all. Therefore, in order to limit the potential for bias, consideration of specific features was not included in the method. However, it is instructive to briefly review what the time-series tend to look like.

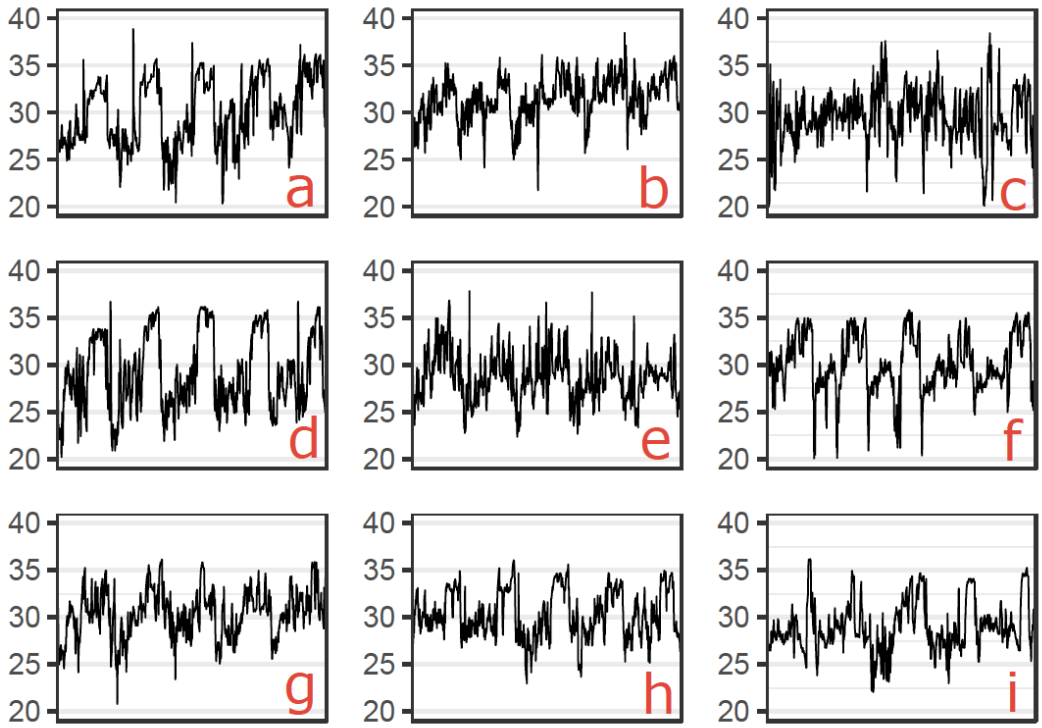


Figure 4.4: A sample of nine randomly selected participant temperature time-series.

Nine randomly selected temperature time-series are shown in figure 4.4. Each shows the 5-day period included in the analysis. Quasi-periodicity is evident in all time-series with variation evident around a 5-day period, with the exception of c, for which the signal is noise dominated. Panels a, d and f exhibit particularly clear diurnal variation, with the warmer periods associated with night-time. Panels a and d have distinct isolated spikes which may correspond to periods of bathing or showering, which is discussed more in the chapter 5 in section 5.4. However, beyond this is impossible to determine what a particular fluctuation corresponds to. It is therefore not possible to estimate the impact of local atypical temperature fluctuations on the metrics. If a participant were to have more showers than average, for example, it is likely that the AX3 would record higher temperature readings. Following this, a higher value of t^m_{10} would be calculated. Therefore, such fluctuations would no longer be atypical because the participant would have indeed experienced a higher temperature than a participant who did not take as many warm showers. A similar argument was made regarding clothing levels in chapter 3. Further consideration of the specific features of the time-series will be the subject of a follow up study (see section 10.2) for which more detailed information will be collected on the activities of daily living which correspond to particular temperature patterns.

4.5 External temperature

The external temperature for the week during which the AX3 was worn was given by gridded NASA MEERA-2 data for daily surface temperature, averaged over the week. Each participant's approximate home location was matched to the corresponding grid square and the 5-day average temperature calculated. The grid resolution in the NASA MERRA-2 dataset is $0.625° \times 0.5°$ which corresponds to approximately 70×35km, which means around 200 grid squares cover the UK (GMAO, 2015, 2016).

Developing this measure of external temperature was not straight-forward. The home location data that the UK Biobank uses is the Ordnance Survey National Grid (OSGB) reference, rounded to the nearest kilometre. The NASA MERRA-2 dataset uses latitude and longitude °values. These two co-ordinate systems are not straightforwardly convertible because of the curvature of the Earth. Therefore, a script was written using a mixture of available Python modules and new code to convert between each system. The two co-ordinate systems are shown in 4.5. This allowed the home location data in the OSGB system to be linked to the specific grid square in the NASA MERRA-2 system. For the particular week that the AX3 was worn the external temperature was taken for the particular grid square that their home was in. The script then used the readings for to external temperature to calculate an average over the week. This average was then inputted into the R database along with the other UK Biobank variables.

4.6 UK Biobank explanatory variables

In addition to the variables of temperature and activity which were derived from the AX3, the following variables were included from the set of possible UK Biobank variables. A brief description is given, including the exact wording of the question presented to participants

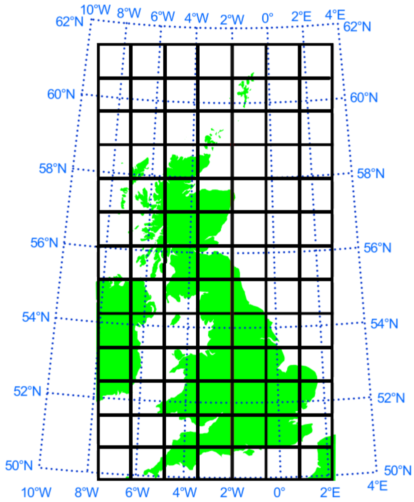


Figure 4.5: The OS grid system (in black) and the latitude longitude grid (in blue). Wikimedia contributor attribution: cmglee, Strebe, MansLaughter, Alexrk2 from naturalearthdata, Pethrus and nandhp `https://commons.wikimedia.org/wiki/File:Ordnance_Survey_National_Grid.svg` licence: `https://creativecommons.org/licenses/by-sa/3.0/legalcode`

during data collection for each variable. Variables were recorded in three *instances*. The initial assessment visit took place between 2006 and 2010. The first repeat assessment (instance 2) was carried out between 2012 and 2013, on a much smaller subset of participants. The most recent assessment was taken at the imaging visit. These first took place in 2014 and are still ongoing (as of July 2019). For participants who attended more than one session, the most recent response was used in this study. The fact that some participant's data is more recent than others is unlikely to impact the results since no participant group's data is systemically more recent than others.

Following the specification of the PAP, discussion with colleagues prompted a further review of the literature that showed was essential that three further variables be included. First, the body mass index (BMI) is an important general measure of health. Given that conditions associated with excess winter deaths are not solely associated with cold exposure, including BMI controls, albeit imperfectly, for the general health of the participant. Moreover, obesity is found to be associated with modified variation in wrist temperature (Corbalán-Tutau et al., 2011). Second, the exclusion of time points in the metric for experienced temperature for which the participant was active only controls for high activity within the t_{10}^{m} variable itself, as a proxy for whether the participant was likely indoors. Given that research question 3.4 uses conditions associated with excess winter deaths (C_{EWD}) as its outcome variable, the activity level recorded by the AX3 needed to be included as an explicit co-variate. Third, the variable of *health satisfaction* was available for a subset of participants (the same subset as financial situation satisfaction and heating type). Given this is a measure of health, albeit subjective, it was determined that its inclusion would be very useful for answering the research questions.

For the variables *fuel type* and *heating Type* participants were able to select more than one option. For these variables, the choices were combined into composite subcategories, as described below.

For each variable, the UK Biobank identification code is given in parenthesis, followed by the official variable name and finally the short title, if required, in square brackets which is used in the results section which follows this chapter. The question that participants saw is also given, including any capitalisation that was used for emphasis.

(54) Assessment centre [centre]. During the initial stages of data collection, participants were invited to one of 21 assessment centres located throughout Britain to gather baseline characteristics. Recruited participants lived within 25 miles of one of these centres at the time of recruitment Fry et al. (2017).

- Barts
- Birmingham
- Bristol
- Bury
- Cardiff
- Cheadle (revisit)
- Croydon
- Edinburgh
- Glasgow
- Hounslow
- Leeds
- Liverpool
- Manchester
- Middlesborough
- Newcastle
- Nottingham
- Oxford
- Reading
- Sheffield
- Stockport (pilot)
- Stoke
- Swansea
- Wrexham

(31) Sex. A binary coding of participant sex, acquired from the central UK Biobank registry.

(20074, 20075) East and north rounded co-ordinate of home location at assessment + NASA data [External temperature]. The home location, which is a rounded to 1km resolution was combined with NASA data, as described above, to provide the approximate average external temperature at the time of wearing the AX3.

(34, 52) Year and month of birth [Age]. These variables were combined to compute the approximate age of each participant, accurate to a month, at the time of wearing the AX3.

(6139) Gas or solid-fuel cooking/heating [Fuel type]. Using the touchscreen interface, participants were asked “Do you have any of the following in your home? (You can select more than one answer)”. The following options were available:

- A gas hob or gas cooker
- A gas fire that you use regularly in winter time
- An open solid fuel fire that you use regularly in winter time
- A gas hob or gas cooker & a gas fire that you use regularly in winter time
- A gas hob or gas cooker & an open solid fuel fire that you use regularly in winter time
- A gas fire that you use regularly in winter time & an open solid fuel fire that you use regularly in winter time
- A gas hob or gas cooker & a gas fire that you use regularly in winter time & an open solid fuel fire that you use regularly in winter time
- None of the above
- Do not know
- Prefer not to answer

(6140) Heating type in home [Heating type]. Using the touchscreen interface, participants were asked "How is your home mainly heated? (You can select more than one answer)". The following options were available:

- Gas central heating
- Electric storage heaters
- Oil (kerosene) central heating
- Portable gas or paraffin heaters
- Solid fuel central heating
- Open fire without central heating
- Two heating types
- Three heating types
- None of the above
- Do not know
- Prefer not to answer

(670) Type of accommodation lived in [Accommodation type]. Using the touchscreen interface, participants were asked "What type of accommodation do you live in?" The following options were available:

- A house or bungalow
- A flat, maisonette or apartment
- Mobile or temporary structure (i.e. caravan)
- Sheltered accommodation
- Care home
- None of the above
- Prefer not to answer

(680) Own or rent accommodation lived in [Tenure type]. Using the touchscreen interface, participants were asked "Do you own or rent the accommodation that you live in?". The following options were available:

- Own outright (by you or someone in your household)
- Own with a mortgage
- Rent - from local authority, local council, housing association
- Rent - from private landlord or letting agency
- Pay part rent and part mortgage (shared ownership)
- Live in accommodation rent free
- None of the above
- Prefer not to answer

(738) Average total household income before tax [Household income]. Using the touchscreen interface, participants were asked "What is the average total income before tax received by your HOUSEHOLD?". Using the help feature on the touchscreen gave participants the weekly and monthly equivalents. The following options were available:

- Less than 18,000
- 18,000 to 30,999
- 31,000 to 51,999
- 52,000 to 100,000
- Greater than 100,000
- Do not know
- Prefer not to answer

During the development of the PAP it was expected that the household income levels could have been adjusted so that, for example, a household with six occupants earning between £18,000 and £30,999 would be assigned a lower per occupant income than the same household earning over £100,000. However, it was later discovered that it is not possible to do this calculation without introducing as many systematic errors as were hoped that the procedure would solve. The two variables were therefore left separate.

(709) Number in household [Household size]. Participants responded to the question "Including yourself, how many people are living together in your household? (Include those who usually live in the house such as students living away from home during term, partners in the armed forces or professions such as pilots)". An integer between 1 and 100 was recorded. Options for 'Do not know' and 'Prefer not to answer' were also included. In order to simplify the analysis, this integer was converted to the following categories: 1, 2, 3 and 4+.

(6142) Current Employment status [Employment status]. The participants were asked to select "Which of the following describes your current situation?"

- In paid employment or self-employed
- Retired
- Looking after home and/or family
- Unable to work because of sickness or disability
- Unemployed
- Doing unpaid or voluntary work
- Full or part-time student
- None of the above
- Prefer not to answer

(4581) Financial situation satisfaction Participants were asked "In general how satisfied are you with your FINANCIAL SITUATION?". The following options were available:

- Extremely happy
- Very happy
- Moderately happy
- Moderately unhappy
- Very unhappy
- Extremely unhappy
- Do not know
- Prefer not to answer

It was expected that *financial situation satisfaction* could provide a proxy for how easily participants were able to afford energy. If those who are unhappy with their financial state were found to have lower experienced temperatures this could be indicative of heat rationing.

(21000) Ethnic background [Ethnicity]. Participants were asked to categorise their ethnic background based on a series of sequential branching questions. This provided granular detail which distinguished between, for example, Black Caribbean and Black African. However, this level of granular detail was deemed unnecessary, so the responses were grouped under the following top-level categories, which corresponds to the first of the sequential branching questions that participants were given.

- Asian or Asian British
- Black or Black British
- Chinese
- Mixed
- Other ethnic group
- White
- Do not know
- Prefer not to answer

(4548) Health satisfaction [Health satisfaction]. Participants were asked to "In general how satisfied are you with your HEALTH?". The following options were available:

- Extremely happy
- Very happy
- Moderately happy
- Moderately unhappy
- Very unhappy
- Extremely unhappy
- Do not know
- Prefer not to answer

(21001) Body mass index [BMI]. This variable was constructed from the weight and height measurements which were recorded during the initial assessment visit. If either of these values was missing, BMI was recorded as NA. The numerical value of the BMI was factored into the standard brackets (Eby and Colditz, 2008) as follows:

- Underweight (BMI < 18.5)
- Normal (BMI 18.5 to < 25))
- Overweight (BMI 25.0 to < 30)
- Obese (BMI > 30))

(41280) Conditions associated with excess winter deaths (C_{EWD}). Conditions associated with excess winter deaths (C_{EWD}) was given a value of 1 if a participant had received a diagnosis of a condition associated with any of the three leading causes of excess winter deaths. These are highlighted by the ONS (2018) report into excess winter deaths. As has been highlighted earlier, the C_{EWD} conditions are circulatory diseases (ICD-10 codes I00 to I99), respiratory diseases (ICD-10 codes J00 to J99). A value of 0 was given to all other participants. Participants with Alzheimer's disease and dementia will be discussed further below in section 4.8.

4.6.0.1 Mean activity level [activity].

The key aspects of the activity level variable were discussed in section 4.3.1.1. Where it appears in the regression models, it is categorised into quintiles. The lowest recorded 5^{th} of the average activity readings appear in the 1^{st} quintile (percentile 0-20), quintile 2 corresponds to percentile $21 - 40$, and so on. The decision was made to divide the activity data this way to allow for the possibility of a non-linear effect as activity increases. In fact, such an approach does not presuppose a monotonically varying relationship at all, and could account for a U-shaped relationship between activity and experienced temperature.

4.7 Multilevel modelling

Multilevel Modelling (MLM) is a method of regression analysis which allows grouped or clustered data to be accounted for (Diez-Roux, 2000; Finch et al., 2014). One common use is in the education system. Typically, student's performance is modelled and compared while taking account of the nested group structure of students within classes, schools and local authorities. Using standard multiple regression models in this context would result in the miscalculation of the errors associated with estimates. In the context of this study, the clearest potential group structure corresponds to the regional centres that the participants attended when the baseline sociodemographic and health data were recorded. These centres were given in section 4.6. Using MLM allows the differences in experienced temperature as a function of sociodemographics to be understood between the different regional centres. Had the study design been one that made use of the time-series data without being summarised as a metric, the MLM structure would have had individual time samples at the lowest level, nested in individuals and then nested in regions.

4.7.1 Statistical analysis

The method for making use of MLM for this study was informed by the tutorials provided by Bristol University's Centre for Multilevel Modelling (Szmaragd and Leckie, 2011). Before describing the regression equations that model the multilevel structure of the regional UK Biobank centres, the initial tests to determine whether multilevel structure exists are described. First, a *null* multilevel model is constructed. For an outcome variable which depends on individuals labelled i nested in groups labelled j, this has the following form

$$y_{ij} = \beta_0 + u_{0j} + e_{ij} \tag{4.1}$$

In this study the outcome y is the experienced temperature, and the groups j each UK Biobank centre. β_0 is the overall mean across all groups and u_{0j} the effect of each group. The final term e_{ij} is the residual error of the regression.

This null model is then compared to a single level model which has no multilevel term u_{0j}, and has the following form,

$$y_{ij} = \beta_0 + e_{ij}. \tag{4.2}$$

In order to determine if there is a significant difference between these two models, the Likelihood Ratio (LR) is calculated using the following equation:

$$LR = -2(\ln(A) - \ln(B)) \tag{4.3}$$

where A and B are the likelihoods of models A and B (the multilevel and single level models in this case). The 5% point of a chi-squared distribution on 1 d.f. is 3.841 and the 1% point is 6.635. Therefore, a value of LR which exceeds 6.635 will provide evidence that there is a significance difference between the two models, and therefore suggests that multilevel structure exists in the data. Finally, in order to determine the magnitude of the variance that can be attributed to this multilevel structure, the variance partition coefficient is calculated.

$$VPC = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_e^2} \quad (4.4)$$

where σ_{u0}^2 is the variance attributed to between group differences, σ_e^2 is the variance attributed to within group differences. Equation 4.4 is a ratio of the variance due to differences between groups to the total variance. Although there is no limit as to what size of VPC should be interpreted as meaningful, a VPC of 0.1 is typical for education studies which analyse the impact of school level differences in attainment Szmaragd and Leckie (2011).

4.7.2 Regression equations

In the circumstance where there is sufficient evidence for multilevel structure, the regression equations take the form described in this section, with variables given the subscript k according to the following list.

4.7.2.1 Variables

The following table summarises the regression variables used in the primary regression.

Variable numbers	Term	Description
20074 + 20075 + NASA	x_1	Mean external temperature (°C)
90004 + processing	x_{14}	Mean activity level (mg)
34 + 52	x_2	Age in years
31	x_3.	Sex (binary)
738	x_6	Household income (categorical)
20119	x_7	Employment status (categorical)
21000	x_9	Ethnic background (categorical)
21001	x_{13}	BMI (categorical)
670	x_4	Accommodation type (categorical)
680	x_5	Tenure type tenure (categorical)
709	x_6	Number in household (categorical)
6139	x_8	Fuel type(categorical)
6140	x_{10}	Heating type (categorical)
4581	x_{11}	Financial situation satisfaction (categorical)
4548	x_{12}	Heat satisfaction(categorical)

Table 4.3: The regression variables used in the primary regression. Three variables were added to those prespecified in the PAP, and are denoted as variables x_{12}, x_{13} and x_{14} (mean activity over the study period). The first grouping corresponds to sociodemographic variables, the second to housing variables and the third to additional variables which were only available for a smaller sub-sample of participants (see 4.6)

4.7.2.2 Regression equations for Research question 1

"Does experienced temperature vary with sociodemographic and building variables [e.g. sex, age, ethnicity, income, building type, tenure]?"

Multiple regression is the appropriate method for estimating the associations between t_{10}^m and other variables. Given the PAP (appendix E) required multilevel structure to be analysed, the regression equations require multilevel structure to be present. The following equation allows for this structure:

$$t^m_{10_{ij}} = \beta_0 + u_{0j} + \sum_{k=2}^{11} (\beta_k + u_{kj}) x_{k_{ij}} \times x_{1_{ij}} + e_{ij} \tag{4.5}$$

where $t^m_{10_{ij}}$ is the experienced temperature of the i^{th} participant in the j^{th} regional centre. e_{ij} denotes the error term, which quantify the offset between the estimate and the measured value of the experienced temperature for a particular participant. The variables run through the k index, as outlined above. The potential of an interaction between each variable and external temperature is provided by the cross term. In practice there may be insufficient statistical power to estimate interactions. All variables are assumed to be linear; this assumption, as well as the others which underpin multiple regression will be examined in chapter 9. Finally, since the number of participants in this study is very high, the significance level for determining whether estimates are statistically significant is set at 99% (H. Kim, 2015) for all regressions in the study.

4.7.2.3 Regression equations for Research question 2

"Are there associations between experienced temperature and the health conditions related to excess winter deaths, i.e. cardiovascular and respiratory diseases, Alzheimer's disease and dementia. [ICD 10 J00J99 F01, F03, G30]?"

As described above, the variable C_{EWD} is a binomial variable which captures whether or not a participant has been diagnosed with a condition associated with EWDs. One of the key requirements for standard multiple regression is that the residuals are normally distributed. A binomial outcome variable will not have normally distributed residuals, and so a generalisation of standard multiple regression is needed. Before describing this, some essential concepts are reviewed (Bonita et al., 2006).

The odds of an outcome are defined as the ratio of the probability of an outcome occurring to the probability of it not occurring, or $\frac{p}{1-p}$, for an event with probability p. For simple cases, OR is estimated from a 2×2 frequency table of exposure and outcome. An example might count the number of people exposed to some level of radiation, and the number of people who develop cancer. The odds ratio would be calculated as follows

$$OR = \frac{a/c}{b/d} \tag{4.6}$$

a is the number people exposed to the radiation who develop cancer.

b the number who are exposed and do not develop cancer.

c the number of people not exposed who develop cancer.

d the number who are not exposed and do not develop cancer.

For $OR = 1$ exposure has no effect on the odds of the outcome. $OR > 1$ is associated with higher odds of the outcome, and $OR < 1$ are associated with lower odds of the outcome. A closely related concept is that of the risk ratio (RR), which is defined as the ratio of the probability of the outcome in an exposed group to the probability of the outcome in an unexposed group.

$$RR = \frac{a/(a+b)}{c/(c+d)} \tag{4.7}$$

Again, a value of $RR = 1$ suggests no effect on the risk of the outcome, $RR > 1$ is associated with an increased risk, and $RR < 1$ a decreased risk. For this second research question, the outcome variable is now C_{EWD}, which is a binomial variable, taking the values either 0 or 1.

As mentioned above, standard multiple regression is not appropriate for binary outcome variables. In order to carry out a regression with such variables, a Generalised Linear Model (GLM) (McCullagh and Nelder, 1989) is required, which allows for non-normal error distributions. The first choice to make when producing a GLM is what kind of error distribution for the outcome variable is appropriate. Since this research question, and the one that follows consists of binomial outcome variables, a binomial error distribution is appropriate.

Following this, a second choice is required. A GLM consists of three components; the systematic component η (the co-variates given by $\beta_k x_k$ in the equation 4.5 for example), the random components μ (the error term e in equation 4.5) and a *link function*, which is a function between the first two components $\eta = g(\mu)$. In standard regression the function g is simply 1. In general, the only requirement on g is that it varies smoothly between its extremes, so a large number of choices of link functions are available. Once the link function is chosen, the regression is performed in a similar manner to simple least squares, where the deviations between the regression line and the data are minimised. In certain cases this cannot be done analytically like in simple least squares, but has to be estimated using numerical techniques. OR is estimated using a logit link function, which is defined as $\ln(\frac{\mu}{1-\mu})$ and sometimes called the log-odds. The RR is estimated using a log link function. Once the parameters are estimated using regression, they must be exponentiated to give the numerical values of OR and RR. Since the form of the regression equations is the same for estimating both OR and RR, the symbol $\mathcal{L}$ is used to denote both the logit and log link function below.

The regression equation for research question 2 tests associations between the variable

C_{EWD} and the experienced temperature y of the i^{th} participant.

$$\mathcal{L}(C_{EWD_{ij}}) = \beta_0 + y_{ij} + e_{ij} \tag{4.8}$$

In practical terms this is implemented in R using the glm function, which carries out generalised linear regression. For outcomes which are rare, OR and RR take the same value, but when the outcome is not rare (around 10% prevalence) the OR is larger than the RR (Davies et al., 1998). Since the outcome variable C_{EWD} is not rare in the data, the risk ratio and odds ratio will have different values. Since RR is much easier to interpret intuitively than OR it will be used in the remainder of this thesis. Whereas as an RR of 0.5 can be interpreted as the risk of an outcome being cut in half, an OR of 0.5 must be interpreted as an odds reduction of 0.5, which is not as straightforward to understand.

4.7.2.4 Regression equations for Research question 3

"Do combinations of sociodemographic factors, building factors and the health conditions related to excess winter deaths (as above) have associations with low experienced temperature?"

This final regression equations combines the variables used in research questions 3.4 and 3.4. The variable C_{EWD} now takes subscripts i and j to denote the i^{th} participant in the j^{th} regional centre, as above.

$$\mathcal{L}(C_{EWD_{ij}}) = \beta_0 + u_{0j} + (\beta_1 + u_{1j})x_{1_{ij}} + \sum_{k=3}^{10}(\beta_k + u_{kj})x_{k_{ij}} \times x_{2_{ij}} + e_{ij} \tag{4.9}$$

These regression equations are converted to single level regression equations by dropping the subscript j. As with the regression equation 4.8, the parameter estimates given by equation 4.9 need to be exponentiated to give the estimates of RR.

4.8 Participant exclusion procedure

The first section of this chapter outlined the process of determining the measure of experienced temperature. The second described the multiple regression models which are used to understand associations between variables. In the final section which follows, the process by which participants are included or excluded in the study is described.

The UK Biobank invited 236,519 participants to wear the AX3 wristband for one week (Doherty et al., 2017). Of these, 103,707 successfully returned the wristband. 102,342 CWA files were successfully processed, and after downsampling to 1-minute period, 101,801 remained. Imposing a minimum wear-time of 90% (in line with Doherty et al. (2017)) reduced the number of participants to 80,046.

The next participant exclusion phase was more complex. The decision was made during the research design stage that participants with either conditions that cause unusually cold hands directly, or those which disrupt circadian rhythms and therefore impact wrist temperature, be excluded. Based on the literature review, the following conditions were identified; Dementia and Alzheimer's disease, anaemia, carpal tunnel syndrome and Raynaud's disease. There are two ways that illnesses are recorded in the UK Biobank database, the self–reported conditions variable and the ICD-10 diagnosis variable. It was decided that self–reporting should be the main way of determining exclusions, since it was important that the condition be sufficiently developed that a participant be aware of it. The exception to this rule of thumb is dementia and Alzheimer's, which may cause cognitive impairment. Alongside these conditions, participants who carried out night shift-work were also excluded, since the literature review identified that the wrist temperature of workers who regularly change their sleeping time is impacted (Jang et al., 2017; Bracci et al., 2016; Ferreira et al., 2013). The impact of this decision is estimated in section 6.3.2. A total of 1,487 participants were excluded on the basis of the above criteria – the number qualifying under each condition is given in table 4.4.

Variable numbers	Description	Number
20002	Dementia/Alzheimer's/cognitive impairment	2
20002	Anaemia	262
20002	Carpal tunnel syndrome	32
20002	Raynaud's phenomenon/disease	47
3426	Night shift (always and usually)	1702
41202	Dementia (including F02 and F03)	6
41202	Alzheimers	4
41202	Raynaud's disease	20
	Total	2081
	Unique Total	2072

Table 4.4: The number of participants who with the given medical condition or who carried out night shift-work

During the course of the study 6 participants decided to no longer participate. Their data was immediately deleted from the database. 118 participants were removed because of data problems, as described in section 4.4. Finally, 49 participants were removed because they had extreme average activity readings ($> 0.1g$), these readings were 8 standard deviations from the mean, and therefore most likely the result of sensor error. This left 78,403 participants in the study. Not all data variables were available for each of these participants. Since multiple linear regression requires complete cases only to be included the number of participants in each model is less than 78,403 - the number of missing or NA values for each explanatory variable is given in table 4.5

Variable numbers	Number of NAs
Average external temperature	0
Sex	0
Age	0
Accommodation type	30
Tenure type	119
Income	291
Employment	27
Household size	263
Fuel Type	206
Ethnic background	83
BMI	157
Average activity level	0
Health satisfaction	40,189
Financial situation satisfaction	40,189
Heating type	40,446

Table 4.5: The number missing or NA values for each variable. The horizontal line divides the two regression models, as *health satisfaction*, *financial situation satisfaction* and *heating type* had similar numbers of missing values

Chapter 5

Results 1: Pilot Studies

...north winds which sadden the most beautiful days produce exactly the effect of those puffs of cold air which enter a warm room through the cracks of a badly fitting door or window.

Victor Hugo – Les Misérables (1862)

5.1 Introduction

This chapter considers the following preliminary research questions, which were first laid out in chapter 3:

P1. Does the AX3 accurately record temperature?

P2. To what extent does it record the immediate thermal environment of the wearer while being worn?

In order to answer these questions, a series of pilot tests were carried out. For each test, the rationale for each test is given, and the results which were expected prior to carrying out the test described. First, a calibration test against a HOBO room temperature monitor is described. Second, a comparison to the approach used by Kuras et al. (2015) is outlined and third, a ten-day trial to understand the device's functionality in real world conditions. Finally, subsidiary pilot studies are described.

Before describing the pilot studies, it is important to highlight the discovery of an error in the original NU processing script. At the start of the pilot studies, three alternative scripts existed for processing the CWA files produced by the AX3. The first of these was a Windows program which was also created by Dan Jackson of Newcastle University, called the OmGUI (`https://github.com/digitalinteraction/openmovement/wiki/AX3-GUI`). This did not provide a way of converting the CWA into a temperature time-series. The second script type was written in MATLAB, and the third in a mixture of Python and Java. Investigations revealed that the output of these two scripts did not match. Following an extensive process to understand each part of the code of both of the scripts, it was discovered that the calibration coefficients included in the Python/Java script did not match the of the MATLAB script. After a series of tests, it was established that the Python/Java script contained an error. Subsequent email exchanges with Dan Jackson of Newcastle University, who developed the software for the AX3, confirmed that the software had been written with the incorrect calibration, and that the author's modification to this was correct. The following code snippet gives the line (1) which was substituted for the correct line (2) in the down-sampling script.

```
doubletemperature=(getUnsignedShort(buf,20)*150.0- 20500)/1000;
doubletemperature=(getUnsignedShort(buf,20)*75.0- 12800)/256;
```

The UK Biobank website published the incorrect average value of the temperature time-series as 20.8°C, since it was calculated using the script which included the error. The correct average temperature reading across all devices is 30.26°C. The process of determining whether the AX3 accurately measures temperature is the topic of the next section.

5.1.1 Calibration study

In order to answer the first of the research questions described above, a calibration study was carried out. The manufacturer's specification indicates that the MCP9700 thermistor used in the AX3 is typically accurate up to ±1°C under the conditions expected for this study with a resolution of 0.3°C (Axivity, 2015). Full specifications were given in section 3.2. Under a wider range of temperature (i.e. above 40°) the inaccuracy can be as much as ±4°C. In order to test these specifications, two unworn AX3 monitors were placed in a small climate chamber. The temperature recorded was compared to a standard calibrated HOBO monitor. The temperature in the chamber was varied between just above 0°C and 40°C over 16 hours, as shown in figure 5.1. This range was selected as it was deemed very likely larger than the possible range of experienced temperatures. The expectation for this study was that the AX3 device readings would agree with the HOBO readings to within the accuracy set out by the manufacture's specification.

A linear regression of the resultant temperature time-series was conducted. AX3 sensor 1 was found to accord with the temperature recorded by the HOBO device with the following regression equation.

$$T_{AX3:1} = 0.997 T_{HOBO} + 0.289 \tag{5.1}$$

with $R^2 = 0.9998$. The second device was found to have the following relationship with the HOBO monitor

$$T_{AX3:2} = 0.991 T_{HOBO} - 0.191 \tag{5.2}$$

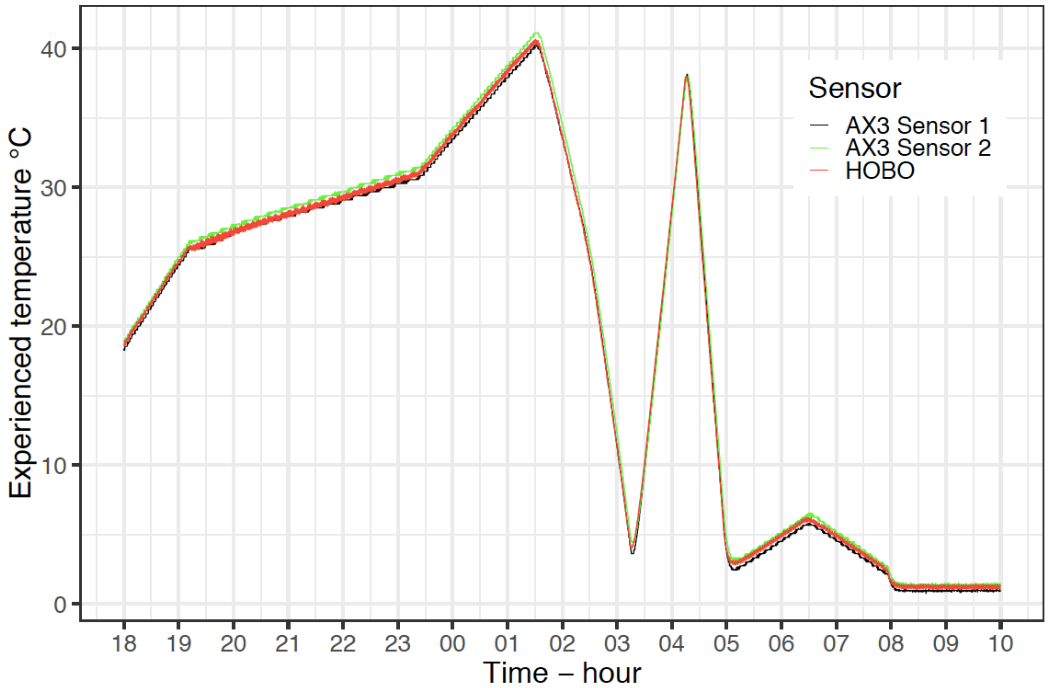


Figure 5.1: The temperature recorded by two AX3 sensors, along with a calibrated HOBO monitor over 16 hours over night in a climate chamber

with $R^2 = 0.9994$. Both devices register different off-sets from the reference HOBO device consistent with the manufacturer's specification. Given that such offsets are not systematically related to the participant classes this study aims to compare, this is not a problem (see section 4.2.1.1).

The thermal response time is a measure of how quickly a sensor responds to temperature changes (Meyer et al., 2008). Typically these are determined under controlled laboratory settings. Since the AX3 sensor reading is determined by ambient temperature as well as other heat gradients, such as incident radiation, the thermal response time will vary depending on environmental conditions. Since the time-series of the AX3 follow the HOBO device, as shown in figure 5.1, the sensitivity to changes is sufficient for this study.

5.1.2 The impact of heat from the wrist

The next pilot test sought to quantify the impact of heat from the wrist on the readings of the AX3. The expectation here was that the AX3 would be influenced by heat from the wrist, but it was unclear to what extent. In order to understand this, the output of the AX3 was compared to two other measurements. The first measured the wrist temperature of the wearer directly by using an iButton temperature monitor directly in contact with the wrist, and secured with medical tape. The second aimed to insulate a sensor worn at the wrist from wrist heat to measure local ambient temperature. For this, a 4cm layer of SpaceTherm Aerogel blanket insulation was placed over the first sensor, taking care to

Figure 5.2: The iButton used to recreate the method of Kuras et al. (2015).

ensure that it remained wrapped in its protective packaging. On top of this, another iButton was attached. Finally, the AX3 was worn alongside this *sensor sandwich*. In the following results summary, parentheses denote the maximum and minimum temperatures recorded. In the office environment of mean 25.4°C (24.4 –25.3), the mean AX3 temperature was 30.1°C (30.0 – 30.2). Wrist temperature was 34.1°C (33.9 –34.2) and the insulated iButton 26.2°C (26.0 –26.3).

The second phase of this test involved going outside into an ambient temperature of approximately 12°C for 5 minutes. During this period the WT dropped by 1.5°C. Both the temperature readings on AX3 and the insulated iButton dropped sharply, by 7.3°C and 7.9°C respectively. The minimum reading on the insulated iButton was 18.4°C and the on AX3 was 22.8°C. Together, these results were taken as evidence that the AX3 is capable of capturing changes in ambient temperature. The thermal response time of the iButton is given as 'up to 130 seconds' by the manufactures. During the period in ambient conditions of 12°C the AX3 registered a drop in temperature at the same rate as the iButton, which suggests the thermal response time is a similar order as the iButton. It should be emphasised that these pilot studies were conducted with an AX3 that was not covered by clothing; an increase of around 2°C was measured when it was covered by a light sweater (it would also be expected that the thermal response time would be longer for a device covered by clothing). Upon returning to the office environment, both the insulated iButton and the AX3 took around 6 minutes to return to the value recorded previously.

5.1.3 Comparison to the method of Kuras et al. (2015)

The next portion of the pilot studies sought to address the second question outlined above, namely to understand the extent to which the temperature recorded by the AX3 reflects the immediate thermal environment of the wearer. As the conceptual model in chapter 3 laid out, the exact definition of the 'immediate thermal environment' is difficult define precisely because it depends on more than just the ambient temperature – incident radiation, moisture levels and wind speed are all important factors for determining heat flow (Gagge and Nishi, 2011). However, the expectation here was that the AX3 would respond to ambient temperature in such a way as to read lower values in cold ambient conditions and higher values in warmer conditions. The counter-factual situation is worth highlighting in this instance. It is perfectly plausible that human thermophysiological responses could be so strong as to mask the impact of changing thermal environments, that is to say, that vasodilatory responses could aim to keep the wrist temperature and its immediate thermal environment at a constant temperature. However, there is no evidence to suggest this is a reasonable expectation, so this portion of the pilot studies was carried out with the expectation that the AX3 would be able to distinguish cold environments from warm ones.

To this end, the next study compared the output of the method used by Kuras et al. (2015), for which an iButton temperature monitor on a fob was worn on the belt (see figure 5.2), with that of the AX3. The set-up was as follows: the AX3 was worn on the wrist, not covered by clothing. An iButton was attached to the belt-loop. Both devices were worn by the author during an afternoon of everyday activity, initially starting out in an office environment before spending time outside in ambient conditions around 9°C, entering a library around 1 hour later, and then returning to the office. The relationship is shown in 5.3. Under stable ambient temperature in the office, the difference between the Axivity and iButton methods was approximately 4°C. The relationship between the iButton and AX3 is not characterised by a simple offset, but generally they agree in form. The origin of the differences observed are likely to be a combination of heat from the wrist which the AX3 is sensitive to, and the smaller amount of heat from the body which the iButton device picks up, as well as different sensitivities to incident radiation. A small transient period is visible at the start of the AX3 time-series where the device warms up due to the proximity of the wrist.

This study showed that the AX3 was able to detect the changes in the thermal environment of the wearer in a way that was commensurate with the iButton.

5.1.4 A group test

A follow-up experiment was conducted for which three colleagues wore an AX3 for the duration of a meeting. The aim of this experiment was to see if AX3 temperature readings were different when worn by different people within the same room. The expectation was that heterogeneity would be observed. The result of this study was that each device recorded different temperatures (each were within 4°C of one another) which varied very little during the course of the 1.5 hour meeting. The room had a perceptible temperature gradient towards the window. Those who were sat closest to the window registered lower

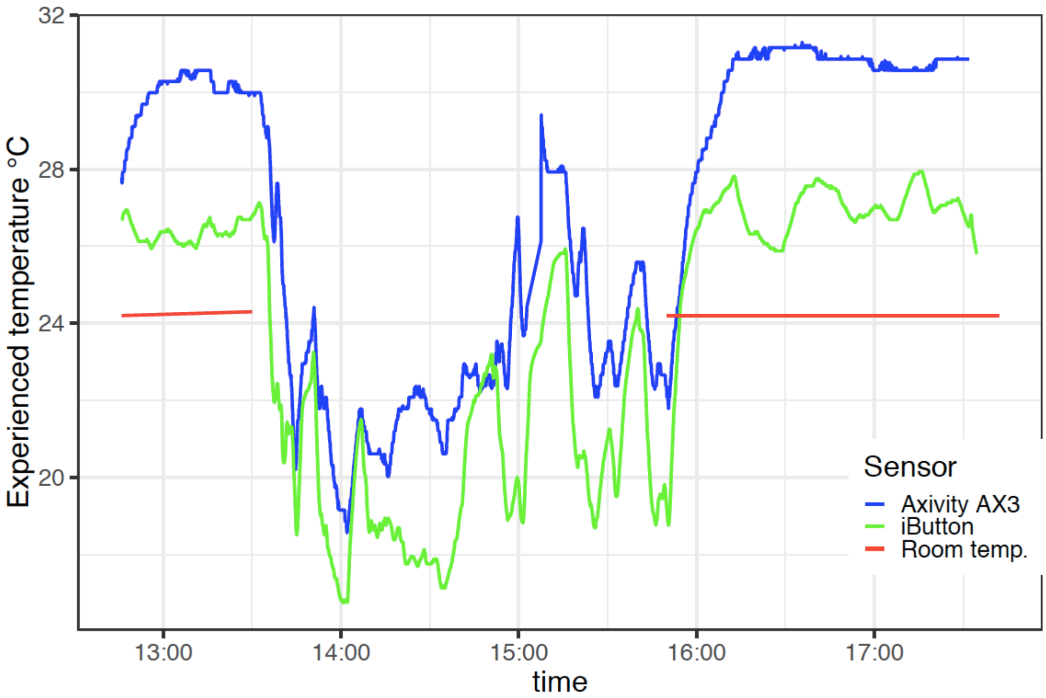


Figure 5.3: The experienced temperature recorded by the AX3 and iButton. The wearer left the office at approximately 13:30 and went outside, before entering a library just after 15:00, going outside and then returning to the office just before 16:00. A transient period at the start of the experiment is visible in the AX3 time-series.

temperatures. There was also an increase in the temperature reading on one device when the wearer decided to roll down their sleeve over the device. However, it was not possible to determine what exactly individual differences in the AX3 readings corresponded to. Large temperature heterogeneity has been observed within rooms, particularly near ceilings – (for example Kane (2013) monitored a single room with 27 different monitors and found substantial temperature variation within the room). It was concluded that the observed variation results from a combination of different local temperature environments in the room as well as differences in how tightly each participant wore the AX3.

5.1.5 Ten-day trial

In order to understand how the AX3 performs under conditions of everyday use, a volunteer consented to wear the device for 10 days. This is three days longer than the wear time used in the UK Biobank, and double amount of time used by this study. The temperature output of this study is shown in 5.4. Temperature peaks at night-time are visible, which correspond to both the increase in wrist temperature while sleeping as well as the warm micro-climate that a bed provides. Indeed, for a simple model of hygrothermal conditions in the bed, (Pretlove et al., 2005) suggests a comfort temperature in bed of 34°C. The night–time temperatures recorded by the AX3 are in the region of 34-36°C, each followed by an early morning sharp peak which corresponded to showering (according to the volunteer). During the day time, the experienced temperature is lower than at night and occasionally features readings lower than 20°C, which likely correspond to periods spent outside.

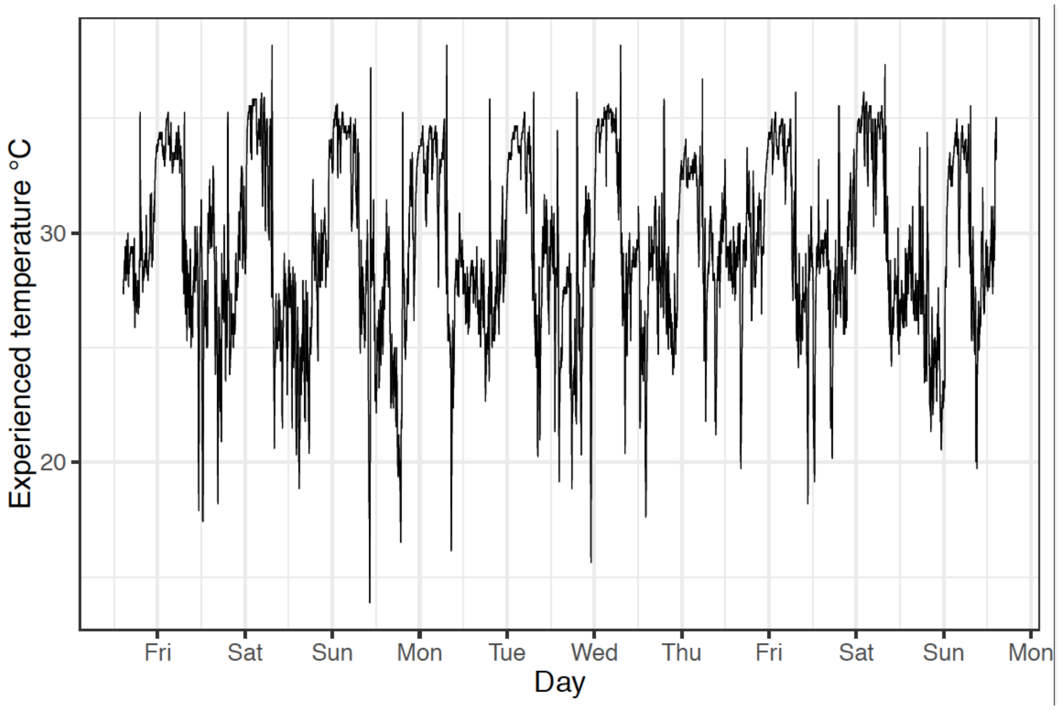


Figure 5.4: The results of a 10-day trial wearing the AX3. Experienced temperature readings vary between 14°C and 38°C

Figure 5.5 show a histogram of the temperature readings for the 10-day trial period. As with the time-series data, a distinct peak is observable at 34°C to 35°C, which corresponds to sleeping periods. The frequency of cold exposure decreases with decreasing temperature, away from the modal reading of 27°C. The wearer experienced a lower frequency of low temperatures and moderate temperatures. It is important to reiterate that these readings do not correspond to the room or environmental temperature directly, but rather the temperature of the on-board thermistor in the AX3 device, which itself reflects ambient conditions.

5.1.6 Light meter

It was initially hoped that the on-board light meter, which purported to measure lux levels, would be of benefit in determining whether the wearer of the AX3 was in bright sunlight. In order to test this, the device was worn during three conditions: bright day time conditions, standard office lighting and in a dark room. Unfortunately neither AX3 test devices registered consistent differences between these conditions. It was therefore concluded that the lux meter is insufficiently accurate to be used in the study.

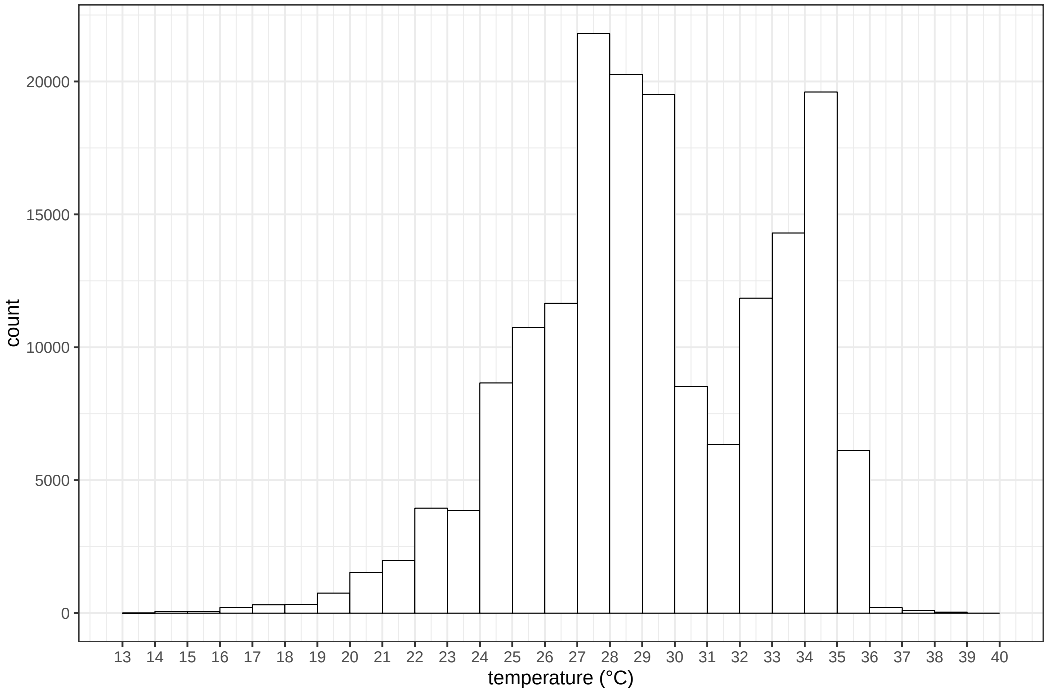


Figure 5.5: The frequency of different experienced temperatures. Two peaks are evident, the peak on the right is associated with sleeping and the central peak to waking periods, most likely spent indoors.

5.1.7 Conclusion

The answer to the research questions for this chapter are as follows. The tests showed that the AX3 measures temperature with sufficient accuracy to justify undertaking the research outlined in the chapter 4. No evidence was found that the AX3 operates outside the tolerances defined by the manufacturers – an individual device measures temperature within ± 1°C of a reference calibrated HOBO monitor.

The second research question of this chapter, regarding the extent to which it records the immediate thermal environment of the wearer is more difficult to answer in a quantitative manner, since a reference device is more difficult to define. However, the tests showed that the AX3 meets the requirement that the reading reflect the ambient temperature. The test using the sensor sandwich suggested the AX3 records a temperature around 5°C higher than room temperature. This finding must be caveated by the fact that the AX3 was not covered by clothing during this test. The impact of clothing on a given reading is one of the greatest sources of uncertainty in this study. However, as was discussed in the conceptual model given in chapter 3, a low reading on the device almost certainly corresponds to the wearer being in a cold environment, even though the converse may not be the case. A full program of study would be necessary to fully determine the impact of clothing on the AX3 readings, which is outside the scope of this thesis. This concludes the first results chapter.

The next chapter explores the relationship between the summary metrics of the experienced temperature introduced in chapter 4, and the sociodemographic factors which were collected as a part of the UK Biobank study.

Chapter 6

Results 2: Experienced Temperature Lower Metrics

I'm cold when the temperatures dip below 70s

Frank Ocean – Biking (2017)

6.1 Introduction

This chapter reports the results of the investigations into the associations between the lower metrics of experienced temperature and the building and sociodemographic factors outlined in the method. Four metrics in total (t^m_{10}, t_{10}, t^m_{min}, t^m_{min}) fall into the category of the lower metrics, which were chosen as they are most likely to capture cold exposure. In order to improve the clarity of the discussion, the first decile metrics (t^m_{10}, t_{10}) are addressed first in full. The results for the minimum (t^m_{min}, t_{min}) metrics then follow. The key findings of this chapter were published in April 2019 in the Journal of Public Health (Kennard et al., 2019). This chapter focuses on answering the first of the core research questions of this thesis, namely "does experienced temperature vary with sociodemographic and building variables [e.g. sex, age, ethnicity, income, building type, tenure]?".

6.2 Space and time

The following section considers the spatial and temporal variation of the recorded experienced temperature. The method outlined in chapter 3 described the multi-level modelling approach, which will be discussed in this section.

6.2.1 Geographical Variation

Understanding whether participants had different experienced temperatures in different parts of Britain is important. It might be expected that different building quality in different parts of the country might lead to different experienced temperatures. It is also plausible that differences in sociodemographics between regions might account for any difference. The regression models used here aim to address this as far as possible with the available variables.

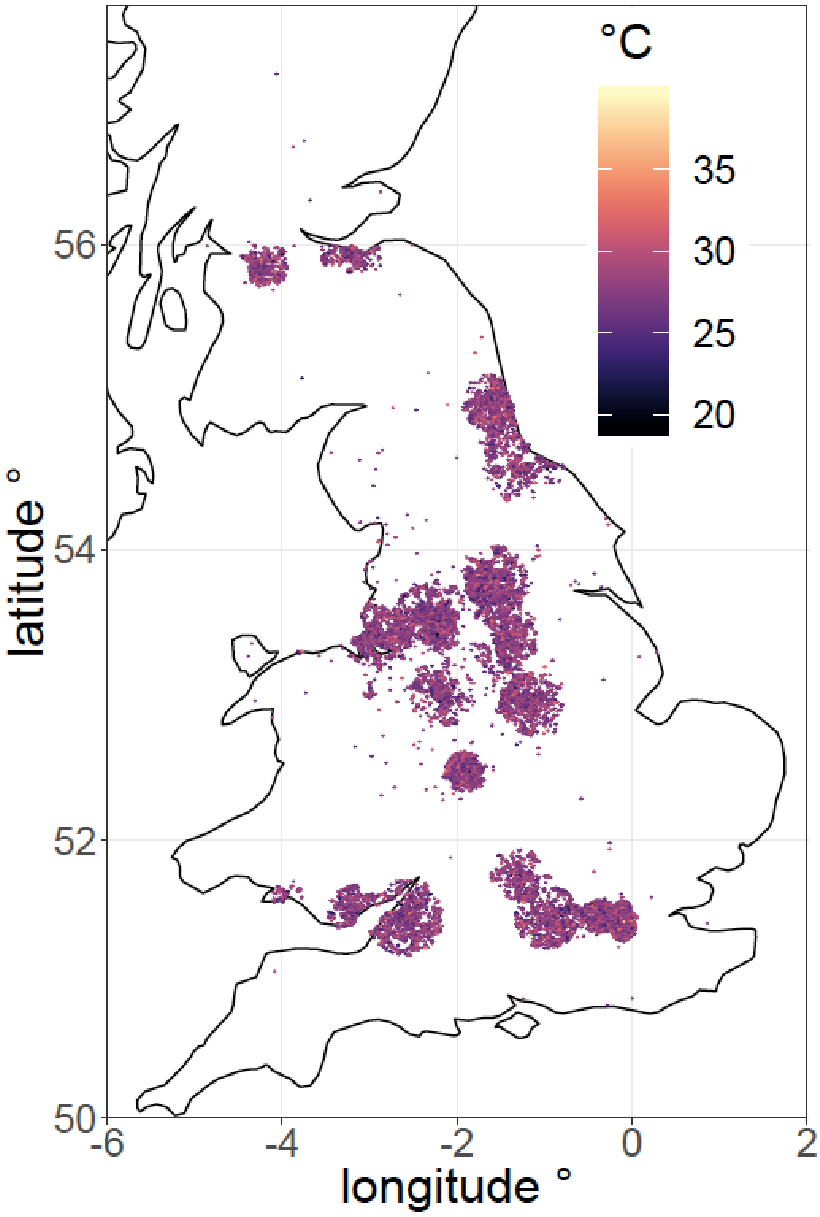


Figure 6.1: The variation in t_{10}^{m} across Britain. Each point corresponds to an individual participant. It is clear that regional variation is minimal. The t_{10}^{m} measurement is seasonally adjusted, as described in the text.

6.2.1.1 Multi-level modelling

In section 4.7 of the methods chapter, the multilevel modelling approach was described as an extension to standard multiple regression which accounts for hierarchical structuring in the dataset. In the UK Biobank dataset, the recruitment centre provides the most natural geographical grouping. The following subsection assesses whether there were significant differences accounted for by these regional centres.

The first measure of difference is the Likelihood Ratio (LR) test (equation 4.3) of the difference between a multilevel model (i.e. one which has regional groupings) and a single level model. This was computed for the whole dataset (N = 77,762) and had the following value:

$$LR = 109 \quad (df = 1) \tag{6.1}$$

There is only 1 degree of freedom difference between the multilevel and single level model, so at the 1% significance level 6.635 is the critical value for the chi-squared distribution. Given that 109 is much greater than 6.635, this is good evidence that multilevel structure exists in the data.

However, the next step requires the magnitude of these differences to be assessed. This was calculated using the Variance Partition Coefficient (VPC) (equation 4.4).

$$VPC = 0.002 \tag{6.2}$$

This value means that 0.2% of the variance in the data can be attributed to regional differences. Since this value is very small, the decision was made not to use a multilevel structure in the analysis of the data. The following results are therefore produced using standard multiple least squares regression. The lack of regional variation in the experienced temperature is shown in figure 6.1, which gives the value of t^m_{10} for each participant at their home location rounded to the nearest kilometer. The seasonal adjustment was achieved by first subtracting the predicted t^m_{10} value obtained from the regression against external temperature, and then adding the mean of t^m_{10}. For t_{10} the values of the VPC was 0.003 and the LR was 192, again indicating good evidence of a negligibly small multilevel effect.

6.2.2 Time of Year

Data collection using the AX3 took place between May 2013 and December 2015. The number of participants monitored in each month of the study is show in figure 6.2, which shows that once the study was established there was good representation of each month during the study period.

6.2.3 External temperature

Figures 6.3 and 6.4 show the relationship between external temperature and experienced temperature for the two metrics discussed in this chapter. Alongside a standard least squares regression line, the locally estimated scatterplot smoothing LOESS regression is included (Jacoby, 2000). LOESS is a generalisation of least squares regression which uses a small region of data - known as the span - to create a local least squares regression. Adjacent regions are then connected together to show if any non-linearities exist in the data. The

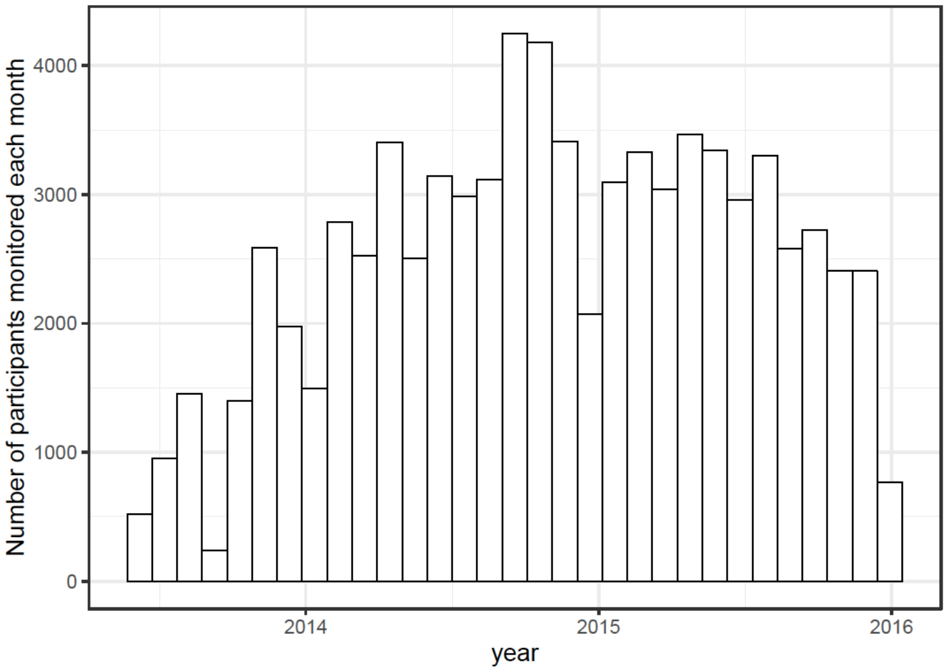


Figure 6.2: The number of participants monitored during each month of the study.

span in the images used here was selected to equate to 1°C average external temperature, which is approximately the accuracy of the measurement. A small deviation from linearity is visible at high temperatures.

Finally, the number of participants monitored at each degree average external temperature is given in figure 6.5. This histogram is equivalent to the average density of figures 6.3 and 6.4 integrated along the vertical experienced temperature axis, and shows that there were fewer readings available at low (< 4°C) and high (> 16°C) external temperature.

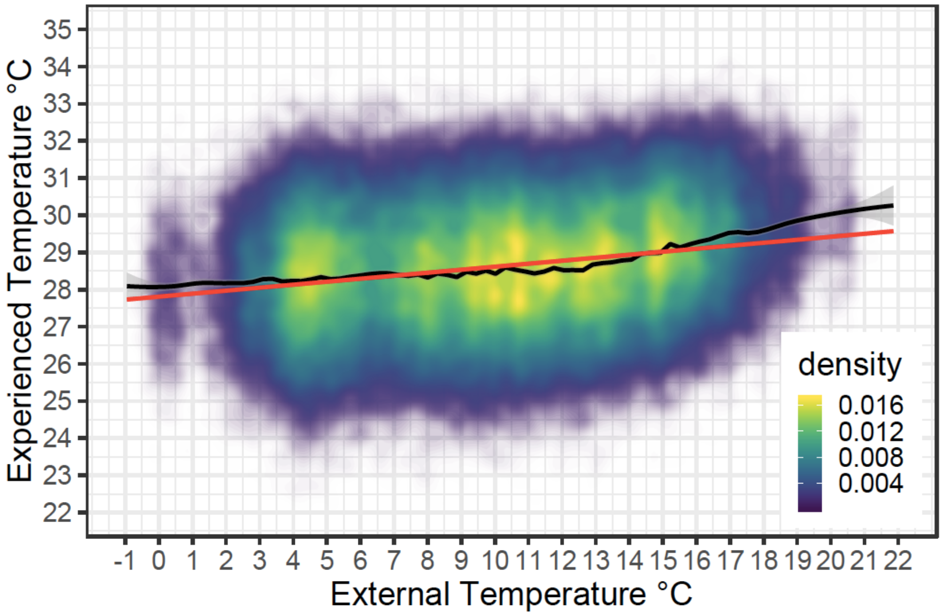


Figure 6.3: The relationship between external temperature and t_{10}^m. The red line is the least squares fit of the data (β 0.08 [0.08 - 0.09], $p < 2 \times 10^{-16}$). The black line is the LOESS regression (see text). Since 77,762 data points are plotted, the data are represented as a density cloud, as given in the key.

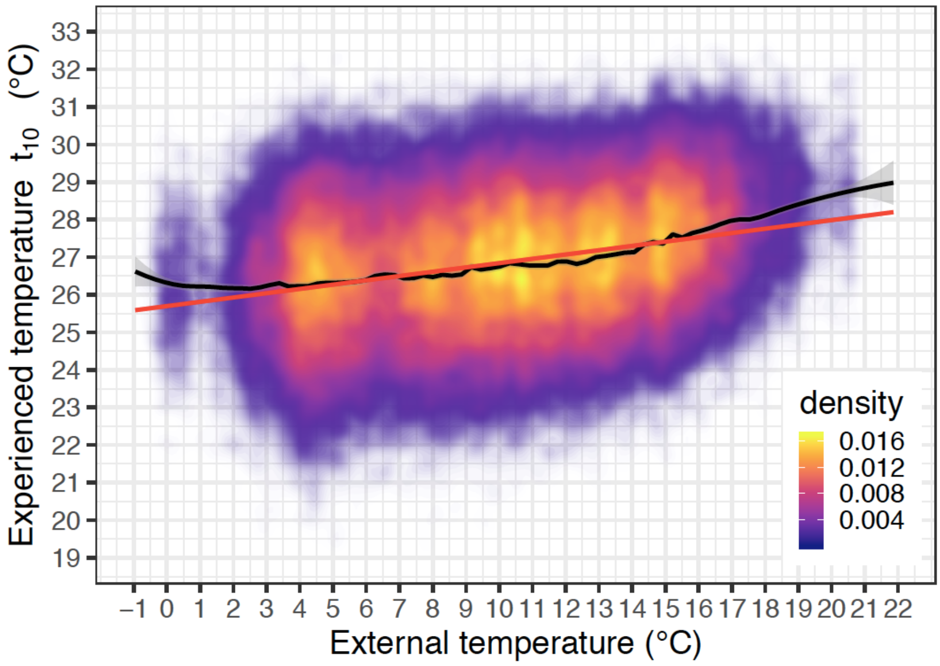


Figure 6.4: The relationship between external temperature and t_{10}. The red line is the least squares fit of the data (gradient 0.12 [0.12 - 0.12], $p < 2 \times 10^{-16}$). The black line is the LOESS regression (see text). 77,762 data points are again represented as a density cloud.

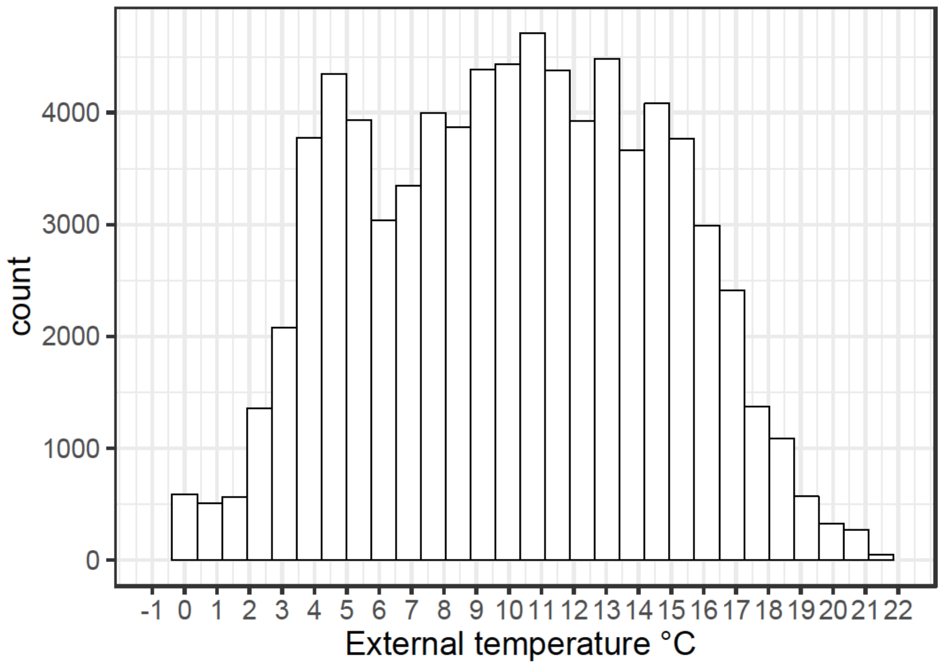


Figure 6.5: The number of participants monitored at each degree average external temperature.

6.3 Sociodemographic and housing factors

This section looks at the combined effect of external temperature and the other variables described in the method. Since multilevel structure is not included, the regression equation tested is simpler than that given in equation 4.5 and has the following form.

$$y_i = \beta_0 + \sum_{k=1}^{11} \beta_k x_{k_i} + e_i \tag{6.3}$$

The index j has been dropped, so that the above equation is for the experienced temperature of the i^{th} participant. The sociodemographic and building variables are given by the index k.

Explanatory variable (relative subcategory, N)	Sub-category (N)	t_{10}	t^{m}_{10}
Intercept	-	26.00 [25.91 – 26.09] ***	27.65 [27.56 – 27.73] ***
External temperature	-	0.12 [0.12 – 0.12] ***	0.08 [0.08 – 0.09] ***
Age (40-49, 6075)	50-59 (21320)	0.06 [0.01 – 0.12]	0.06 [0.01 – 0.11]
	60-69 (35407)	0.16 [0.10 – 0.22] **	0.18 [0.13 – 0.24] ***
	70-79 (14960)	0.33 [0.26 – 0.40] ***	0.28 [0.21 – 0.34] ***
Sex (Female, 43770)	Male (33992)	-0.10 [-0.13 – -0.07] ***	-0.10 [-0.13 – -0.07] ***
Ethnic background (White, 75365)	Mixed (398)	-0.08 [-0.27 – 0.11]	-0.11 [-0.29 – 0.07]
	Asian (654)	0.23 [0.08 – 0.38] *	0.12 [-0.02 – 0.27]
	Black (582)	0.09 [-0.07 – 0.25]	0.11 [-0.04 – 0.26]
	Chinese (157)	-0.36 [-0.66 – -0.06]	-0.62 [-0.90 – -0.33] **
	Other ethnic group (395)	0.00 [-0.19 – 0.19]	-0.11 [-0.29 – 0.07]
	Do not know (20)	0.02 [-0.82 – 0.86]	-0.20 [-1.00 – 0.60]
	Prefer not to answer (191)	0.06 [-0.22 – 0.33]	-0.02 [-0.28 – 0.24]
Household Income £ (Less than 18,000, 10592)	18,000 to 30,999, (17779)	0.06 [0.01 – 0.11]	0.05 [0.01 – 0.10]
	31,000 to 51,999 (20016)	0.03 [-0.02 – 0.07]	0.01 [-0.04 – 0.06]
	52,000 to 100,000 (17021)	0.03 [-0.02 – 0.09]	0.01 [-0.04 – 0.06]
	Greater than 100,000 (4850)	0.06 [-0.01 – 0.13]	0.00 [-0.06 – 0.07]
	Prefer not to answer (5475)	-0.19 [-0.60 – 0.23]	0.08 [0.01 – 0.14]
	Do not know (2029)	-0.52 [-1.60 – 0.55]	0.15 [0.06 – 0.24] *
Accommodation type (House/bungalow, 71554)	Flat (6058)	0.19 [0.13 – 0.24] ***	0.13 [0.07 – 0.18] **
	Temporary (54)	-0.07 [-0.59 – 0.44]	0.07 [-0.42 – 0.55]
	None of above (83)	-0.05 [-0.18 – 0.08]	-0.14 [-0.54 – 0.25]
	Prefer not to answer (13)	-0.17 [-0.51 – 0.17]	-0.36 [-1.39 – 0.66]
Tenure type (Own outright, 44537)	Mortgage (28498)	0.06 [0.02 – 0.09] *	0.01 [-0.02 – 0.04]
	Rent Local Authority (2096)	0.23 [0.14 – 0.32] **	0.12 [0.03 – 0.20] *
	Rent private (1497)	-0.03 [-0.13 – 0.07]	-0.10 [-0.20 – -0.00]
	Shared (174)	-0.01 [-0.30 – 0.27]	-0.03 [-0.31 – 0.24]
	Rent free (469)	0.16 [-0.01 – 0.34]	0.01 [-0.16 – 0.18]
	None of above (276)	0.09 [-0.14 – 0.32]	0.00 [-0.22 – 0.22]
	Prefer not to answer (215)	-0.14 [-0.41 – 0.12]	-0.19 [-0.45 – 0.06]
Household size (single occupant, 12854)	Two (37905)	0.11 [0.07 – 0.15] **	0.19 [0.15 – 0.23] ***
	Three (12141)	0.09 [0.04 – 0.14] **	0.14 [0.09 – 0.18] **
	Four or more (14862)	0.07 [0.01 – 0.12]	0.10 [0.05 – 0.15] **

Table 6.1: The results of the regression model of the associations between the experienced temperature summary metric and the sociodemographic and building variables. The table gives results for using both t_{10} and t^{m}_{10} as outcome variables. N = 77,762 after incomplete cases were removed. Square brackets denote the 95% confidence intervals on the model parameter estimates. Significance levels: * $p < 0.01$, ** $p < 0.001$, *** $p < 1 \times 10^{-9}$.

Explanatory variable (relative subcategory, N)	Sub-category (N)	t_{10}	t^m_{10}
Employment status (In paid employment or self-employed, 39797)	Retired (27472)	-0.05 [-0.08 – -0.01]	0.03 [-0.01 – 0.07]
	Looking after home/family (3235)	-0.04 [-0.11 – 0.03]	-0.04 [-0.10 – 0.03]
	Unable to work, sickness/disability (1411)	0.21 [0.11 – 0.32] **	0.10 [-0.01 – 0.20]
	Unemployed (901)	-0.02 [-0.15 – 0.11]	-0.09 [-0.21 – 0.03]
	Doing unpaid or voluntary work (3759)	-0.07 [-0.13 – -0.00]	-0.03 [-0.09 – 0.03]
	Full/ part-time student (738)	-0.11 [-0.25 – 0.03]	-0.11 [-0.24 – 0.02]
	None of the above (350)	-0.20 [-0.40 – 0.01]	-0.20 [-0.39 – -0.01]
	Prefer not to answer (99)	0.17 [-0.21 – 0.55]	0.19 [-0.17 – 0.56]
Fuel type (Gas hob or gas cooker, 28957)	Gas fire (6379)	-0.02 [-0.07 – 0.03]	0.00 [-0.05 – 0.05]
	Open solid fuel fire (2335)	-0.27 [-0.35 – -0.19] ***	-0.15 [-0.22 – -0.07] **
	Gas hob & Gas fire (20188)	-0.02 [-0.06 – 0.01]	0.00 [-0.03 – 0.03]
	Gas hob & Open fire (4481)	-0.18 [-0.24 – -0.12] **	-0.08 [-0.14 – -0.02] *
	Gas fire & Open fire (195)	-0.45 [-0.72 – -0.18] *	-0.39 [-0.65 – -0.13] *
	Gas hob & Gas fire & Open fire (956)	-0.23 [-0.35 – -0.10] **	-0.22 [-0.33 – -0.10] **
	None of the above (14221)	0.06 [0.02 – 0.09] *	0.06 [0.02 – 0.10] *
	Prefer not to answer (37)	0.50 [-0.14 – 1.13]	0.40 [-0.21 – 1.00]
	Do not know (13)	-0.47 [-1.52 – 0.57]	-0.72 [-1.71 – 0.28]
Body Mass Index (Normal, 30562)	Underweight (477)	-0.11 [-0.28 – 0.06]	-0.17 [-0.33 – -0.00]
	Overweight (45722)	0.15 [0.12 – 0.18] ***	0.05 [0.02 – 0.08] **
	Obese (1001)	0.19 [0.07 – 0.31] *	-0.17 [-0.28 – -0.05] *
Activity level quintile (1st quintile, 15463)	2nd quintile (15567)	-0.40 [-0.44 – -0.36] ***	-0.07 [-0.11 – -0.03] *
	3rd quintile (15567)	-0.69 [-0.73 – -0.64] ***	-0.16 [-0.20 – -0.11] ***
	4th quintile (15578)	-0.97 [-1.01 – -0.92] ***	-0.26 [-0.30 – -0.22] ***
	5th quintile (15587)	-1.44 [-1.48 – -1.39] ***	-0.45 [-0.50 – -0.41] ***

Table 6.2: The continuation of table 6.1.

Explanatory variable (relative subcategory, N)	Sub-category (N)	t_{10}	t^m_{10}
Financial situation satisfaction (Extremely happy, 3808)	Very happy (14498)	0.02 [-0.05 – 0.09]	0.01 [-0.06 – 0.08]
	Moderately happy (15732)	0.01 [-0.06 – 0.09]	0.00 [-0.07 – 0.06]
	Moderately unhappy (2473)	0.05 [-0.05 – 0.16]	0.01 [-0.09 – 0.11]
	Very unhappy (737)	0.20 [0.04 – 0.36]	0.13 [-0.03 – 0.28]
	Extremely unhappy (369)	0.10 [-0.11 – 0.31]	0.03 [-0.17 – 0.24]
	Prefer not to answer (57)	0.03 [-0.47 – 0.54]	-0.26 [-0.74 – 0.22]
	Do not know (56)	0.20 [-0.31 – 0.71]	0.08 [-0.41 – 0.56]
Health satisfaction (Extremely happy, 2230)	Very happy (13771)	0.06 [-0.03 – 0.15]	0.01 [-0.07 – 0.09]
	Moderately happy (17767)	0.14 [0.05 – 0.22] *	0.03 [-0.05 – 0.11]
	Moderately unhappy (2955)	0.20 [0.09 – 0.31] **	0.05 [-0.06 – 0.15]
	Very unhappy (661)	0.27 [0.10 – 0.45] *	0.07 [-0.09 – 0.24]
	Extremely unhappy (249)	0.36 [0.09 – 0.62] *	0.14 [-0.11 – 0.39]
	Prefer not to answer (10)	-0.39 [-1.59 – 0.80]	-0.21 [-1.35 – 0.93]
	Do not know (87)	0.20 [-0.22 – 0.61]	0.14 [-0.26 – 0.53]
Heating type (Gas central heating, 34999)	Electric storage heaters (798)	0.06 [-0.08 – 0.20]	0.10 [-0.03 – 0.24]
	Oil (kerosene) central heating (979)	-0.14 [-0.27 – -0.01]	-0.06 [-0.19 – 0.06]
	Portable gas or paraffin heaters (10)	-0.05 [-1.24 – 1.14]	0.12 [-1.01 – 1.26]
	Solid fuel central heating (128)	-0.43 [-0.76 – -0.09]	-0.14 [-0.46 – 0.19]
	Open fire without central heating (109)	-0.18 [-0.54 – 0.19]	-0.23 [-0.58 – 0.12]
	Three heating types (5)	0.49 [-1.19 – 2.17]	0.15 [-1.45 – 1.75]
	None of the above (676)	-0.07 [-0.22 – 0.08]	-0.07 [-0.21 – 0.08]
	Prefer not to answer (15)	0.78 [-0.33 – 1.88]	0.70 [-0.35 – 1.75]
	Do not know (11)	-0.37 [-1.52 – 0.78]	-0.55 [-1.65 – 0.55]

Table 6.3: The results of the extended regression model, including all variables given in tables 6.1 and 6.2 but with *financial situation satisfaction*, *heating type* and *health satisfaction* added. The table gives results for using both t_{10} and t^m_{10} as outcome variables. N = 37,730 after incomplete cases removed. Square brackets denote the 95% confidence intervals on the model parameter estimates. Significance levels: * $p < 0.01$, ** $p < 0.001$, *** $p < 1 \times 10^{-9}$.

Broadly speaking, t_{10}^m is designed to quantify differences in cold exposure for sedentary or low activity periods of the participant's daily life. These will most likely correspond to times when the participant was indoors, although not exclusively. t_{10} on the other hand uses all activity levels in the calculation of the metric, so will capture cold exposure in general. All the results are given in tables 6.1 and 6.2, but for clarity each variable will be addressed individually. The model revealed a number of significant differences of at least 0.1°C at the $< 1\%$ significance level for both t_{10}^m and t_{10}.

Mean external temperature. For the relationship between external temperature and t_{10}^m, and external temperature and t_{10}, figures 6.3 and 6.4 show regression gradients 0.08 and 0.12 respectively. After the other explanatory variables were included the relationships do not change. This suggests these gradients are robust since they are not influenced by the inclusion of other variables. These estimates can be compared to one given by Chambers and Oreszczyn (2019), who estimated the gradient of the average change in internal domestic temperature with external temperature in the UK to be 0.17.

Age. Relative to those aged 40-49, t_{10}^m was found to be 0.16°C higher for those aged 60-69 and 0.33°C higher for those aged 70-79. A similar result is found for t_{10}, which found increases of 0.18°C and 0.28°C respectively.

Sex For both t_{10}^m and t_{10}, male participants were 0.10°C colder than female participants

Ethnic background. For t_{10}^m, the only significant difference relative to those who identified their ethnic background as White was for those who identified as Chinese - t_{10}^m was found to be 0.62°C colder. A single significant difference was found for t_{10} – Asian participants (excluding Chinese) were 0.23°C warmer than White participants.

Household income. No significant results were found as a function of income level, with the exception of those who reported not knowing their household income. The differences relative to those earning less than £18,000 for this category were 0.13°C and 0.14°C for t_{10}^m and t_{10} respectively.

Accommodation type and tenure type. For accommodation type, only participants who lived in flats had significantly different t_{10}^m and t_{10}, at 0.13°C and 0.19°C warmer respectively. For tenure type, those living in accommodation rented from the Local Authority had t_{10}^m 0.12°C warmer than those who owned their home outright and for t_{10}^m the difference was 0.19°C.

Household size. Almost all categories of household size differed significantly from the reference single occupant household. The greatest effect was for two person households, for whom t_{10}^m and t_{10} were 0.19°C and 0.11°C higher, respectively, than those who lived alone. The differences from single person occupancy became less large as the number in the household increased after this initial level.

Employment status. Those who were unable to work due to sickness or disability were the only participant type found to be significantly different from those in paid or self–employment. The difference for t_{10} was 0.21°C. For t^m_{10} no significant difference was observed.

Fuel type. The key significant difference between the reference category of those who only had a gas fire or gas cooker was the presence of an open solid fuel fire. All sub-categories that included an open solid fuel fire were significantly colder in terms of t^m_{10} and t_{10}.

Body Mass Index. Compared to normal BMI participants, overweight and obese participants had higher t_{10} by 0.15°C and 0.19°C respectively. For t^m_{10} there was a very small but significant difference for overweight participants of 0.05°C. Interestingly, the effect direction reversed for obese participants, whose t^m_{10} was found to be 0.17°C colder than those of normal BMI.

Mean Activity. Higher activity levels were consistently associated with lower t^m_{10} and t_{10}. This was the only variable for which all subcategories were significantly different from the reference.

Additional variables. For the variables of *financial situation satisfaction* and *heating type* there were no significant differences for t^m_{10} or t_{10}. However, for *health satisfaction* t_{10} increased monotonically as a health satisfaction decreased. It is notable that this difference was found even when the effect of activity level was controlled for. The fact that this difference was not observed for t^m_{10} may indicate that those with lower health satisfaction are exposed to less cold temperatures when not sedentary.

In the PAP, the possibility of testing for interactions between explanatory variables was included. It was decided that including these in the models reported here would result in too great a level of complexity, so they are not explicitly considered further.

6.3.1 Statistical checks

The method of multiple regression used in this section relies on the use of the Ordinary Least Squares (OLS) method of fitting a line to data. For a given variable, the process estimates the gradient and intercept of a linear relationship between explanatory and outcome variables by minimising the sum of the square deviations between the data and the line. The use of OLS depends upon several requirements about the data that should be met in order for it to be applied. The following section examines these assumptions in turn, and determines the extent to which the statistical approach used here is mathematically justified given the data.

6.3.1.1 Linearity

The first requirement for linear regression is that all continuous variables, in the case of the model above that is *external temperature*, have a monotonic relationship with the outcome variable. This means that any complex relationship, such as a U-shaped or S-shaped form, should not be fitted with linear regression. Graphical confirmation that the relationship between external temperature and both t_{10}^{m} and t_{10} conforms to this requirement is given in figures 6.3 and 6.4.

6.3.1.2 Residuals

The second requirement is that the residuals – the difference between the predicted and actual values of the regression – are normally distributed. Again, this can be confirmed visually, as is done in figure 6.6.

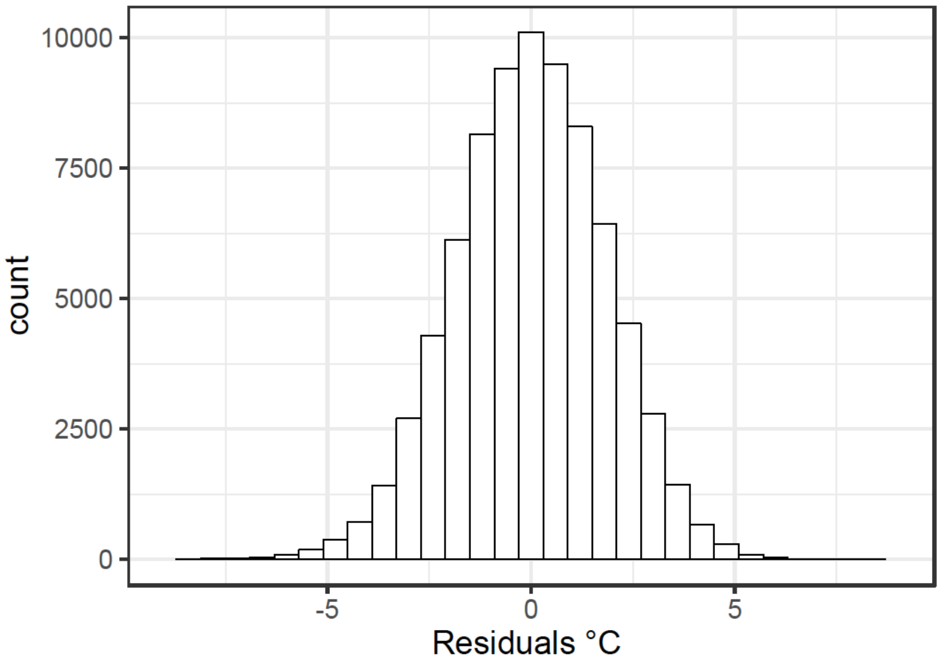


Figure 6.6: The residuals for the above fit of t_{10}^{m}. They are normally distributed. The standard deviation of the residuals 1.83°C

6.3.1.3 Multicollinearity

If explanatory variables in a multiple regression are correlated, the parameter estimates might be incorrect. In order to check for multicollinearity, the Variance Inflation Factor (VIF) is calculated. Values for the VIF of each variable of the model using t_{10}^{m} are given in table 6.4. The table also includes the VIF adjusted for the degrees of freedom (df) of the variables (Fox and Monette, 1992). The critical value at which multiple collinearity is likely a problem is 10. Since all values in table 6.4 are much smaller than 10, multicollinearity is not a problem for this model.

6.3.1.4 Regression dilution

Another important requirement for the correct application of OLS is that the explanatory variables should be measured without error. If this condition is not met, the result is

Variable	VIF	df	$VIF^{1/(2df)}$
External temperature	1.0	1	1.0
Age	2.0	1	1.4
Sex	1.1	1	1.0
Accommodation	1.3	4	1.0
Tenure	1.8	7	1.0
Income	1.5	6	1.0
Household size	1.3	1	1.2
Employment	2.0	8	1.0
Fuel type	1.1	9	1.0
Ethnicity	1.1	7	1.0
BMI	1.1	3	1.0
Activity level	1.2	4	1.0
Heat	2.0	9	1.0
Financial situation satisfaction	1.0	7	1.0
Health satisfaction s	1.4 7	1.0	

Table 6.4: Variance inflation factor (VIF) and VIF adjusted for degrees of freedom (df) to 1 d.p for the model for t_{10}^m

a systematic bias of the estimated gradient towards zero, known as attenuation bias or regression dilution (Kennard et al., 2018). In multiple regression models, this can result in the remaining variance being shifted onto the other variables, so that the effects estimates of those variables are over-estimated. It is therefore important to check whether this is likely to impact the results described in this section. This is carried out by estimating the ratio of the error on the explanatory and outcome variable. In situations where the error on the explanatory variable approaches that of the outcome variable, regression dilution can start to have an impact, although determining specific thresholds is a matter of some discussion in the literature (Berglund, 2012; Hutcheon et al., 2010). However, as the following argument demonstrates, making a specific calculation is unnecessary in this instance. For external temperature, the error can be separated into two separate terms. First, there is the random error associated with whether the NASA MERRA-2 data accurately captures the average external temperature for a given week at a given location. Simple estimates of the error associated with NASA MERRA-2 temperature measurements are not common. However, one study of sea surface temperatures, which are affected by moisture levels and therefore subject to error, reports error estimates. For these challenging conditions, the difference between satellite temperature measurements, from which NASA MERRA-2 is derived, and in-situ temperature measurements have an average error of around ±0.5°C. Land measurement temperature errors are likely to be less than this. The error on an individual AX3 wristband is certainly greater than this value (±1°C), and so the respective ratio is probably not large enough to cause a problem in this case. The second kind of error is harder to quantify, and corresponds to the ambiguity in defining the variable of interest – this is the same question as asking if at the time of measurement was a particular participant's local external temperature well described by the NASA MERRA-2 reading. For example, given that the external temperature reading was derived from the participant's home location, there is a chance they were not in the same location as their home for some portion of the study week. It is possible, for example, that they collected the AX3 device at their home in Scotland, left for France, and then returned before the study week was over. The external temperature difference between these two locations would lead to a significant systematic error in the external temperature reading. This problem is extremely hard to quantify exactly, but it is very unlikely that a significant proportion of the participants were not in the vicinity of their home for the study period. The fact that the device was sent to the participant's home goes some way to guarding against issues in this regard, though it is certainly the case the validity of this variable relies on the assumption that the participant's external temperature is sufficiently well described by the NASA MERRA-2 data.

6.3.2 Robustness

6.3.2.1 Excluded participants with particular conditions

In the method chapter 4.8, the exclusion criteria for participants with particular conditions was outlined. As part of this, it was decided that self-reported conditions associated with particularly cold hands should be excluded. In order to understand the impact of this decision, a robustness check using anaemia was carried out. The various conditions of anaemia were chosen because they had the largest disparity between self-reported and health practitioner diagnosed rates. This involved excluding the additional participants who had ICD-10 codes associated with anaemia. Over 2,000 such participants were excluded, in addition to those already removed. This had almost no impact on the results. Where effect estimates differed it was by no more than 0.02°C, and all were within the confidence intervals of the original dataset which only used self–described exclusions. A complete analysis of the impacts of night-shift work on experienced temperature is one potential avenue for future work, described in section 10.2, but is outside the scope of this study.

6.3.3 First decile

The calculations of the experienced temperature metrics discussed here make use of midpoint interpolation to calculate the first decile in the event that it lies between two values (see appendix C.2 for the code). However, it is also possible to use linear interpolation in this function. Midpoint interpolation takes the mean of two adjacent points, whereas linear interpolation incorporates the gradient of the line between the two points to determine the intermediate value. The difference between linear and midpoint interpolation in the decile computation algorithm was found to account for at most a difference of ±0.01°C within a particular time-series – the final impact of the regression estimates is therefore smaller than this, and is not observed in their estimation. While this difference is small, it is important to note that computational techniques necessarily involve approximation.

6.3.3.1 The median level of activity

The final portion of this section examines the robustness of the methodological choice of using a participant's median activity level as a cut off to divide high activity periods from low activity or sedentary periods. The simplest method of assessing this was to re-run the analysis using different criteria. The most obvious alternative is to use a hard cut-off which is defined for the population in general and not one that varies from person to person. Determining what this cut-off should be challenging. The literature suggests that the cut-off for Moderate to Vigorous Physical Activity (MPVA) is a recording of 100mg on an activity wrist band (Menai et al., 2017). Specifically, in order to be classified as MPVA the authors used a method for which "mean acceleration in the 5s-epoch... needed to be at or above 100mg. To remove signals related to random wrist movement, we only retained activities lasting at least 1 minute for which 80% of the activity satisfied the 100mg threshold criteria." Given the downsampled AX3 data used in this study has a 60-second period, a requirement of 100mg across the entire minute would be stricter than the method used by Menai et al. to classify a 60-second period as MPVA. The lowest average activity level that could be considered MPVA using the Menai et al. method can be calculated by considering the slightly artificial scenario in which activity of 100mg is measured for 80% of the 5-second readings in a minute (10 epochs), and zero otherwise. This would be equal to 83.3mg on average. This threshold of 83.3mg per 60-second period would still tend to include some periods of mild activity. Therefore, a lower threshold of 10mg was chosen as an average over the 60-second period to ensure that periods of MPVA were more likely to be excluded, so that the time periods remaining in the metric would almost certainly be low-activity periods. This limit was denoted as t^f_{10}. As a check, another fixed threshold of 100mg for the 60-second period was also assessed, and given then name t^q_{10}.

Since the motivating questions of this thesis are around energy use in buildings specifically, it is worth reiterating that the goal of the method of excluding activity above the median was to ensure that temperature time periods included in the metric corresponded to times of low activity, and therefore periods for which the participant was more likely than not to be indoors. The 10mg limit was designed to be stricter to exclude high activity behaviour, as opposed to a method which would delineate the boundary between low and moderate activity. This is doubly important since the converse of the above argument does not hold – it is not true to say that high levels of activity necessarily mean an occupant is likely outside, even though it is probably more often the case, all other things being equal.

In order to test if these thresholds produced different results to those screened by the median individual activity levels, the whole dataset was reprocessed, and the regressions described above carried out with both t_{10}^{f} and t_{10}^{q}.

The results showed that the t_{10}^{f} metric produced very similar regression estimates as t_{10}^{m}. The root mean squared difference between the effect estimates for the t_{10}^{m} and t_{10}^{f} regressions was 0.04°C, with a mean difference of -0.01°C. All explanatory variables which were found to be significant for t_{10}^{m} were also significant for t_{10}^{f}. Two further variables for t_{10}^{f} showed statistical significance below the $p < 0.01$ level. Namely, participants who identified their ethnic background as Black had t_{10}^{f} of 0.2°C higher than White participants ($p < 0.01$), (cf. an insignificant difference of 0.12°C for t_{10}^{m} for the same variable). Those who had a mortgage were found to be significantly warmer by 0.04°C than those who owned their own home ($p < 0.01$) for t_{10}^{f}. The same effect estimate was found for t_{10}^{m} but the difference was not significant. Performing an OLS regression of the effect estimates for t_{10}^{m} against t_{10}^{f} found a relationship of gradient $\beta = 0.98$ and intercept $\alpha = 0.0001$ with an $R^2 = 0.96$. Taken together this suggests a very similar outcome using the median activity level at the individual level versus a general threshold of 10mg to examine temperature time periods in which the participant is most likely to be sedentary.

For t_{10}^{q} the outcome is much closer to the t_{10} metric, as might be expected, since fewer periods qualify for exclusion under this less strict threshold and the temperature time-series is closer to the unfiltered case. All but two variables which were found to have significant differences for t_{10} are significant for t_{10}^{q}. The exceptions are that those who identify their ethnic background as Asian are found to be significantly warmer than those who identify as White. Similarly, those who live rent free are also not found to be significantly warmer to those who own their own home. An OLS fit between the regression estimates found a relationship with $R^2 = 0.97$, gradient $\beta = 0.90$ and intercept $\alpha = 0.001$. The root mean squared difference between the coefficients was 0.04°C and the mean difference was 0.01°C. The full list of regression coefficients is given in appendix B. The heterogeneity of the kinds of results described in the previous two paragraphs makes summarising them difficult, but the overall picture is that t_{10}^{f} and t_{10}^{m} are very similar to each other, and t_{10}^{q} and t_{10} are also similar.

Direct comparison between the effect sizes for t_{10}^{m} and t_{10} is not meaningful since they aim to describe different phenomena (however the standardised effect sizes are compared in figure 6.10). t_{10} aims to capture general cold exposure, t_{10}^{m} aims to capture cold exposure which occurs indoors. There are a number of counter examples that can be considered which test the validity of these assumptions, such as times sitting waiting for a bus. In these situations however it is more likely that a participant is wearing a higher level of clothing than when indoors. If certain groups of people were systematically spending a lot of time outdoors, poorly dressed, and at a sedentary level of activity it is expected that such groups would be shown to have significantly different experienced temperatures than the reference group using t_{10}^m – in this specific and somewhat artificial example it would be beneficial that the model identified these groups as being different. The converse situation, in which people are sedentary outdoors during the warmer summer months is more likely, but for these times there is much less chance of cold exposure occurring. Considerations such as these will be examined further in chapter 9.

6.4 Exploring the first decile

The relationship between t_{10}^m and the unfiltered metric t_{10} has not yet been considered in detail. This section examines the relationship between them with a view to justifying alternatives. The relationship between them is given in figure 6.7. The figure shows that the majority of participants which lie along the $X = Y$ line have low activity. Equivalently, the majority of the lighter coloured points, which correspond to the participants that recorded the highest activity during the study week, appear below the $X = Y$ line. In general, the points appearing below the line indicate that having a higher activity level is associated with a lower t_{10} for a given value of t_{10}^m. Given that the average temperature inside UK homes is warmer than the outside for all but the hottest summer days, it can be concluded that the decision to remove higher activity periods does reflect times when the participant was in a warmer environment, which is most likely to correspond to indoor environments. As a comparison, plotting figure 6.7 coloured by external temperature quintile reveals no such pattern.

In assessing whether an improvement upon the t_{10}^m and t_{10} metrics can be made, it is helpful to consider their shortcomings. Firstly, it can be noted that the choice of the 10^{th} percentile was arbitrary. On the one hand, it was chosen as a compromise between the minimum, which was expected to be prone to outliers in the time-series (see section 6.5), and the mean, which was expected to be dominated by warm periods in bed that would not be able to characterise cold exposure. In these regards the tenth percentile functions well. On the other hand, there is no reason why any other lower percentile value would not have functioned as well the tenth. Therefore, it is helpful to consider whether alternative metrics may perform the role of characterising cold exposure effectively without the issue of arbitrariness being present.

6.5 Standardised effect size and the minimum experienced temperature

In this section the results are expanded to include minimum experienced temperature metrics. A plot of the relationship between t_{min} and t_{10} is given in figure 6.8. The two

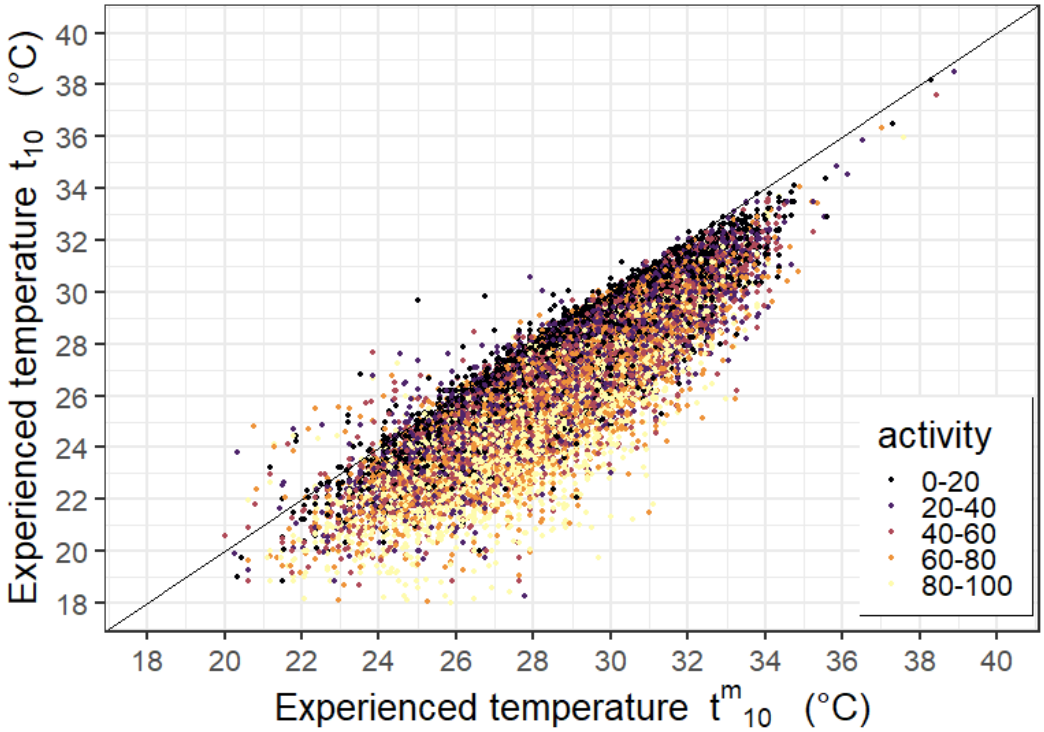


Figure 6.7: Comparing t_{10} with t^m_{10}. Each data point is coloured by the activity level quintile. The $X = Y$ line is given.

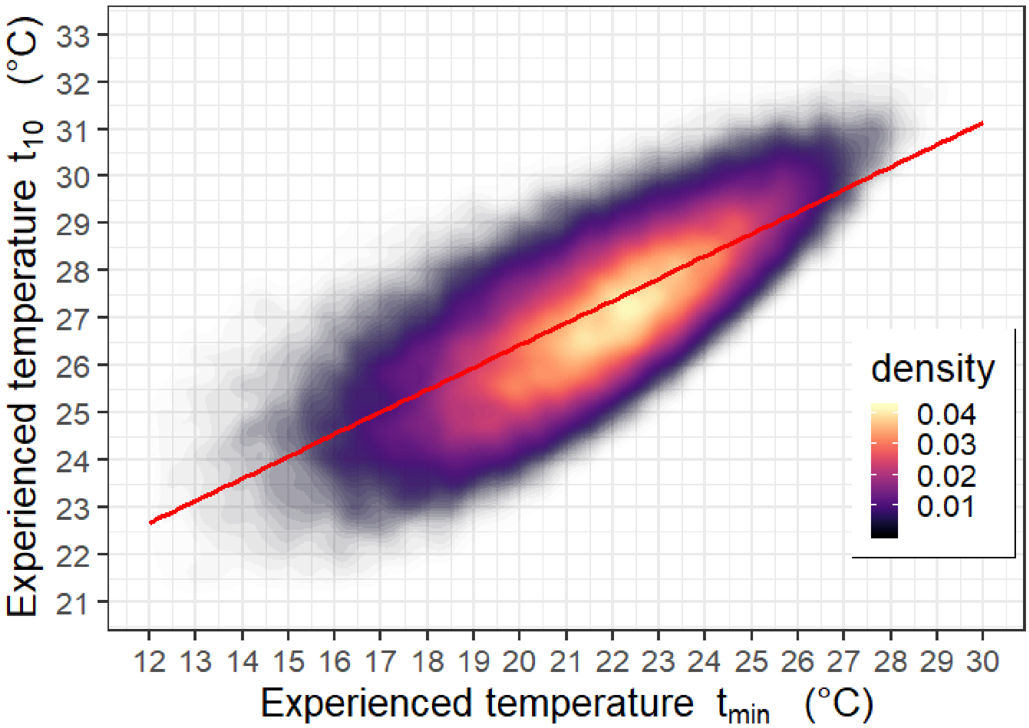


Figure 6.8: The relationship between t_{min} and t_{10}. The red line is the least squares fit of the data $t_{min} = 1.21t_{10} - 11.62$, $R^2 = 0.56$. The data points are represented as a density cloud

metrics are highly correlated, with $R^2 = 0.56$. Before examining the regression results for the minimum experienced temperature (t_{min} and t^m_{min}), it is instructive to consider standardised effect sizes, which will allow a means of comparing the first decile and minimum metrics. It is common to estimate the standardised effect size when reporting the results of quantitative studies. Broadly, there are two classes of standardised effect size. Firstly, the r class includes measures of correlation such as r^2 and Pearson's r. The second class d comprises measures such of standardised difference, such as Cohen's d, Hedge's g, or Glass's Δ. The latter type is typically calculated by normalising effect estimates by the standard deviation. Baguley (2009) points out that simple, unstandardised effect size is often preferable, because the standardisation procedure may be biased by restricted sampling practices, wherein the standard deviation is artificially reduced and the standardised effect size over-estimated. For example, this would occur if outliers were systematically removed from one group and not another. However, given this does not occur in this study, standardisation is a robust means of comparing the metrics used here.

The approach adopted here is to standardise the metrics prior to their inclusion in the regression model. This is mathematically equivalent to calculating Cohen's d (defined as the difference between group means divided by the population standard deviation (Rosnow and Rosenthal, 2003)). As a rule of thumb, Sawilowsky (2009) gives the following levels of interpretation of Cohen's d; $d = 0.01 \implies$ very small, $d = 0.2 \implies$ small, $d = 0.5 \implies$ medium, $d = 0.8 \implies$ large, $d = 1.2 \implies$ very large, and $d = 2.0 \implies$ huge. It is important to stress that these rules of thumb do not define universal equivalence scales - even Cohen himself warned against the blind use of such rules of thumb with respect to the interpretation of d (Cohen, 1988). Much like the choice of p-value provides a means of distinguishing significant from insignificant effects, setting a threshold of standardised effect size, for example $d = 0.1$, gives a way of focusing only on the effects which are large enough to warrant particular attention. Again, there is no a priori reason why $d = 0.1$ should be chosen, but since no other work has been conducted on the relevant effect size of differences in large sample experienced temperature studies, this threshold provides structure.

With these restrictions on interpretation in mind, the lower panel of figure 6.10 gives the effect sizes, denoted $d(x)$, for the significant results (p$<$ 0.01%) for t^m_{10} and t_{10} metrics. The upper portion of the figure gives the comparison standardised effect size values of t^m_{min} and t_{min} . For completeness, the numerical values findings of using t_{min} and t^m_{min} are given in the appendix in tables B.1, B.2 and B.3. The findings for the subcategories which are significantly different to the reference category for this metric are similar to that of t_{10} and t^m_{10}. Unsurprisingly, t_{min} and t^m_{min} are found to increase with average external temperature with comparable effect sizes as for t_{10} and t^m_{10}.

In general, figure 6.10 shows that standardised effect sizes for the significant results in t_{10}

Metric	Summer-winter standardised effect size	Effect size interpretation
t_{min}	0.96	large
t^m_{min}	0.72	medium
t_{10}	0.72	medium
t^m_{10}	0.48	medium

Table 6.5: The estimates for the summer-winter effect size difference of average external temperature

are larger than t^m_{10}. This is likely reflective of the observation that domestic temperatures are generally warmer than external temperatures in the UK, so that sedentary periods, which the t^m_{10} aims to capture, necessarily involved less cold exposure than times where the participant is active.

The effect size for average external temperature per degree is very small for all of the lower metrics, but given a typical 12°C variation in external temperature between summer and winter in the UK, the overall effect size is larger for the summer-winter difference taken together. The overall summer-winter standardised effect size estimates for the impact of average external temperature is given in table 6.5.

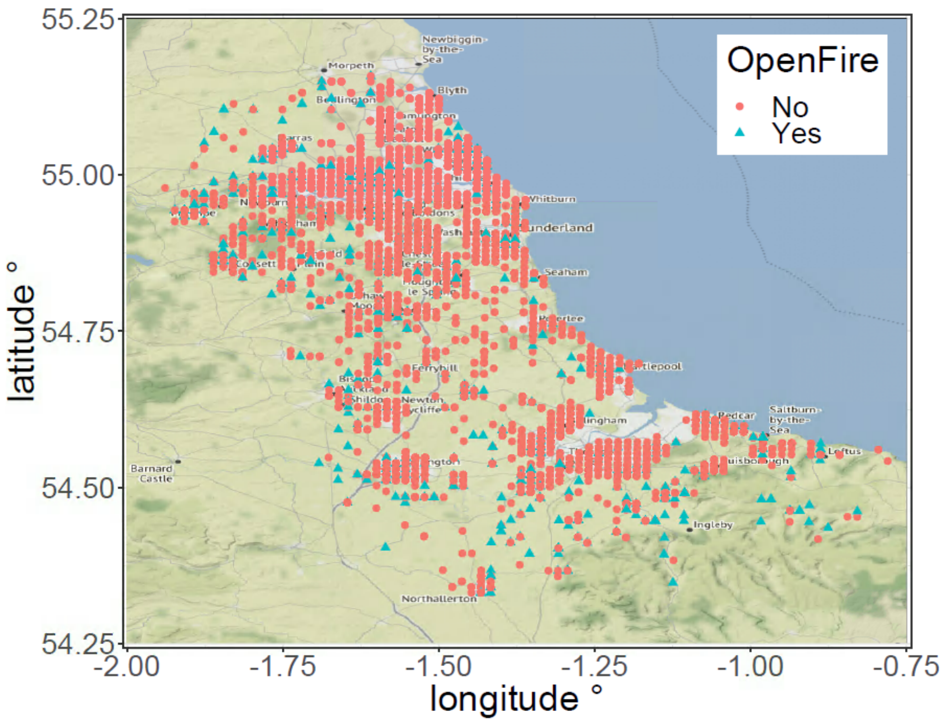


Figure 6.9: The distribution of UK Biobank homes with and without a solid fuel open fire in the North East of England, in the areas near Newcastle and Sunderland. It is evident that those homes with solid fuel open fire are more likely to be located outside the urban areas

There is a consistent effect that living in a house with an open solid fuel fire, whether coupled with other gas heating or cooking appliances, is associated with colder t^m_{10}, t_{10},

t_{min} and t^m_{min}. Several explanations for this are plausible. Open fires tend to heat a much more local area of the home than central heating, so participants may be exposed to lower temperatures in rooms of the house which are not near the open fire. Moreover, homes with open fires tend to be located in more rural areas. This is demonstrated in figure 6.9. Participants living in more rural environments might be exposed to colder temperatures than their more urban counterparts due to an absence of the urban heat island effect, especially if living in older housing stock which is often harder to heat than more modern buildings (Palmer and Cooper, 2013). This possibility would require further in-depth research (see 10.2). The use of the external temperature variable would not be able to account for this difference, since the spatial resolution is unable capture the differences between rural and urban locations.

In terms of ethnic background, the standardised effect estimates across the four lower experienced temperature metrics are inconsistent. The effect that participants of Chinese ethnic background have t^m_{10} effect size of -0.35 relative to participants with a White ethnic background is not significant in other metrics. This may be due to the relatively small sample size for those who identify as having a Chinese ethnic background (n=157). One anecdotal explanation of the effect, if it is a real one, given by colleagues at the UCL Energy Institute is that the use of electric blankets for personal heating while sedentary, which is common in China, might account for the difference. Since the AX3 is worn on the wrist, it would measure ambient air temperature as lower for periods of time where it was not covered by the blanket than in homes that made more use of central heating systems which heat air to provide comfort. Again, such a speculation would require in-depth work to assess. Asian participants (excluding Chinese) have higher t_{min}, t^m_{min} and t_{10} than White participants. Again, the sample size is relatively small (n=654). t_{min} shows significantly higher readings for participants whose ethnic background is Black (n=582). Again, even though significant results are found, they should be treated with a degree of caution as they are difficult to justify theoretically, but could be used as a basis for further investigation.

A similar effect as a function of age is shown for the minimum metrics as to the first decile metrics. Minimum temperatures increase with age, and there is again almost no difference between the unfiltered and median filtered metrics. Like with t_{10}, this likely reflects the uniformity of the thermal experience as a function of whether the participant was active or not and probably that older people spend less time outdoors.

Household size shows consistent but very small differences across the lower experienced temperature metrics. Compared to those living alone, households of size two show consistently warmer experienced temperatures. This is difficult to account for. It could be that homes with multiple occupants have longer heating periods on average because they have to cater to more than one thermal comfort preference. This could reduce the amount of cold periods that the occupant living in a multi-occupant home would be exposed to.

For the sex variable, there is a small effect size difference. Males have a lower minimum temperature than females. This is consistent with the findings for both t_{10} and t_{10}^{m} discussed above.

Unlike the first decile metrics, a clear income effect is evident for the minimum temperature. Both t_{min} and t_{min}^{m} decrease with income level. Those who live in households earning more than £100,000 per year, have t_{min} -0.46 [-0.57 - -0.35]°C and t_{min}^{m} of -0.21 [-0.31 - -0.11]°C, relative to those living in household earning less than £18,000 per year. As figure 6.10 shows these differences correspond to standardised effect sizes of -0.14 and -0.07 respectively. Intermediate income levels follow this trend. This finding will be discussed more in the following chapter that considers the diversity of temperatures experienced using the thermal variety metric.

There is evidence that those who live in Local Authority rented accommodation have higher minimum experienced temperatures than those who own their homes outright, but the difference is not significant for t_{min}^{m}. This accords with the picture presented by t_{10}, although a significant difference was also found using t_{10}^{m}. There are no significant differences for those who live in flats for the minimum temperature metrics.

Body Mass Index results for overweight participants are consistent across the four metrics, each show a slightly increased experienced temperature relative to normal BMI participants. The picture is less clear for the obese participants. While t_{min} and t_{10} are significantly higher for obese participants compared to those with normal BMI, the findings for t_{min}^{m} and t_{10}^{m} are not consistent. The effect for t_{min}^{m} is still that obese participants have higher minimum experienced temperature, but the finding is not significant ($p = 0.02$).

The findings for recorded activity level are consistent with expectations. Minimum and first decile experienced temperatures reduce monotonically with increased activity levels. This result is also seen for t_{10}^{m} and t_{10}.

From the additional variables which were only available for a smaller subset of the data, the clearest findings are for t_{min} and t_{10}. Both metrics show experienced temperatures increases with health dissatisfaction. These results are not significant for t_{10}^{m}, but for t_{min}^{m} there is a similar but smaller effect which is consistent with t_{min} and t_{10}. Overall it is possible to conclude that participants who are more satisfied with their health experience colder temperatures than those who are more dissatisfied. These findings generally accord with the only significant variable as a function of employment status. This finds that those who are unable to work due to sickness of disability are have higher t_{min}, t_{min}^{m} and t_{10} experienced temperatures.

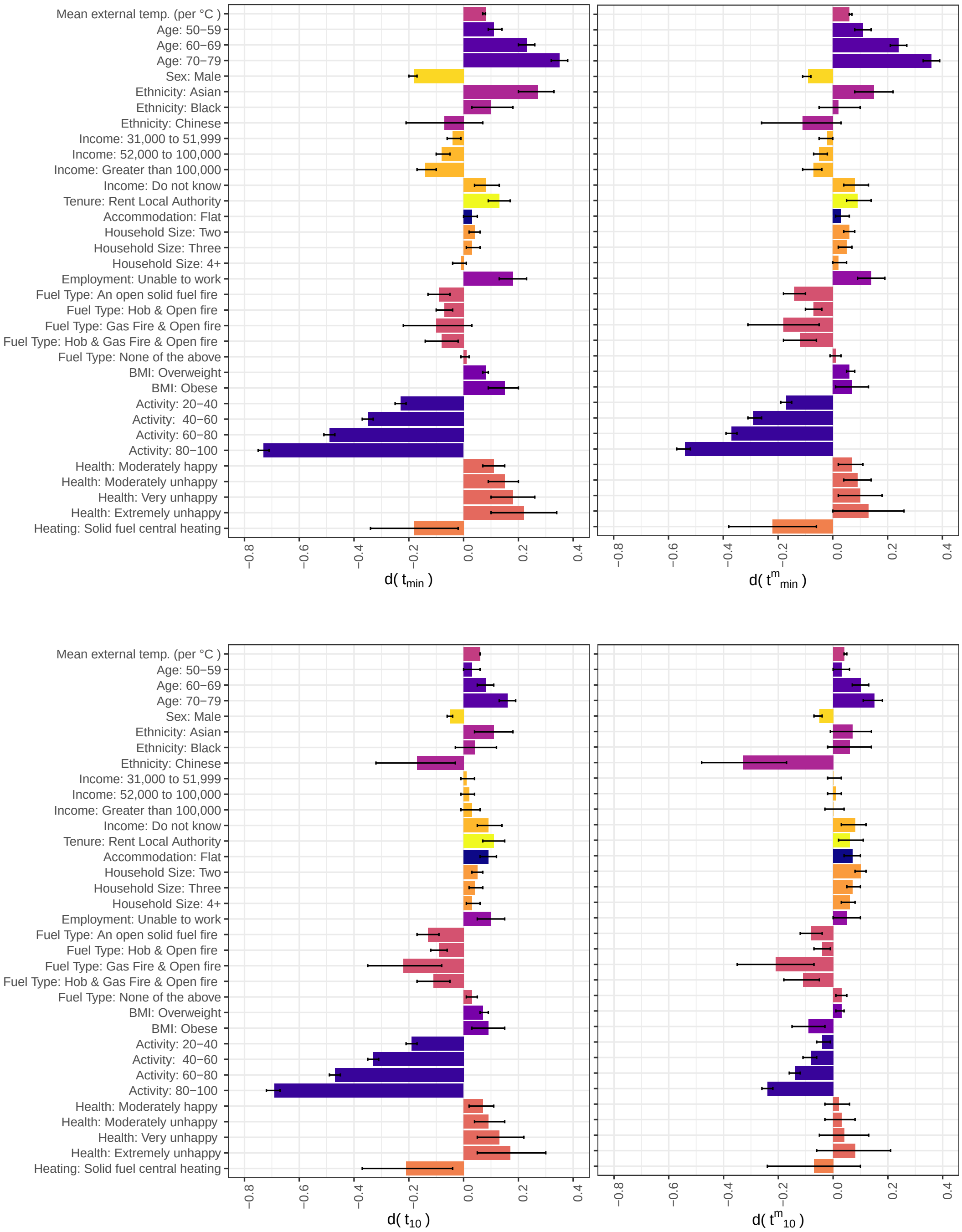


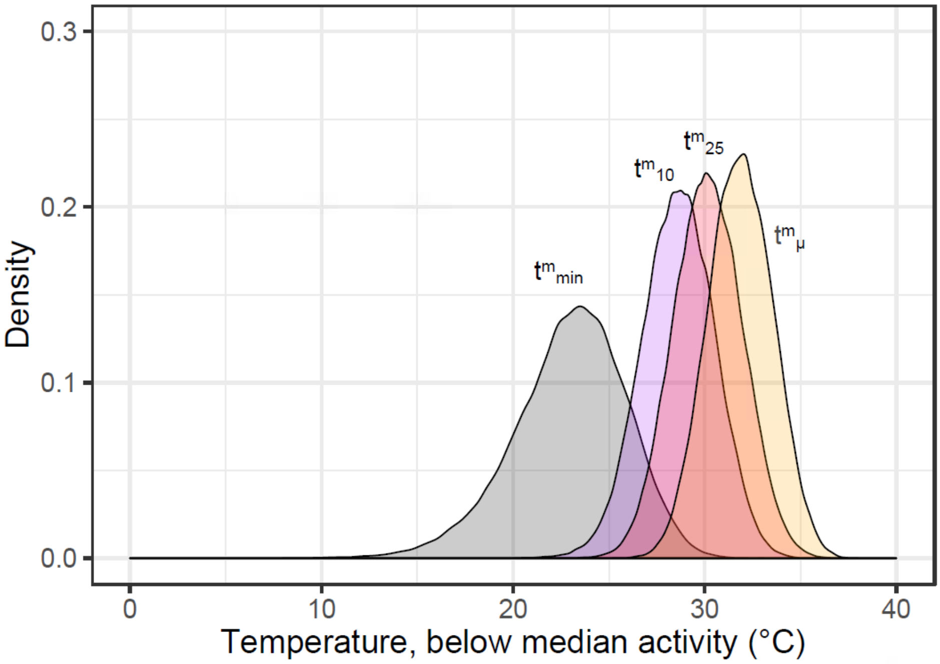
Figure 6.10: Effect size estimates for t_{min}, t^m_{min}, t_{10} and t^m_{10}, for significant subcategories only. The 95% confidence intervals are denoted by error bars. N=77,762 for all variables other than *health* and *heating*, for which N=37,730.

The findings given here suggest the minimum experienced temperature to be a useful compliment to the first decile metric. Generally, but with some noted exceptions, the two pictures agree. The meaning of these findings will be discussed in depth in chapter 9.

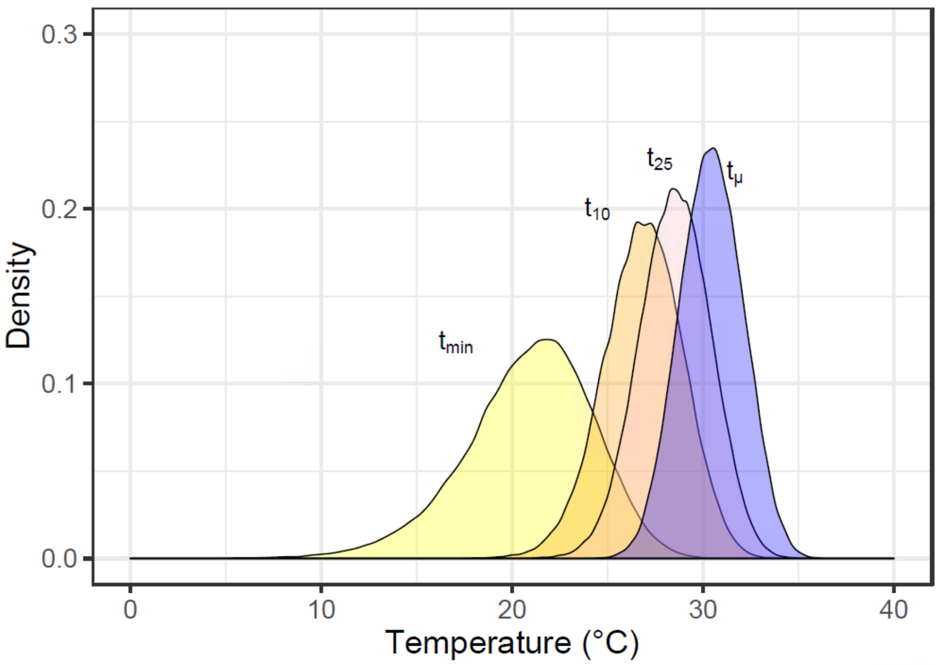
6.6 Alternative metrics

It is useful to consider the distribution of the metrics in general in order to contextualise the above findings. Figure 6.11b gives the distribution of the values of each of the metrics, with the addition of the first quartile t^m_{25}, and its equivalent for the unfiltered time-series, t_{25}. The figure makes clear that the variance in the metrics which use the minimum value is greatest, and the mean shows least variation. The minimum values are generally much further away from the first decile, first quartile and mean measurements than they are from each other. A minimum value reflects a brief sampling of the lowest experienced temperature, whereas the metrics of the first decile and first quartile function as a compromise between the minimum and mean values. However, while it is true to say that for an individual participant the minimum is prone to outliers, this is not true when use the minimum in a regression model. For a regression model, each regression group estimate effectively averages over the individuals in the group. Therefore, if one or even several of the individual minimum readings are themselves outliers, a group estimate in a regression will not be as prone to being affected by outlying values.

Although they are not reflective of cold exposure, the metrics from the upper portion of the distribution are also important (see table B.1). The 3$^{\text{rd}}$ quartile, t_{75} and maximum reading t_{max}, are most likely associated with bathing/showering and the warm micro-climate of the bed, since neither internal, external nor wrist temperatures are associated with temperatures around 37°C (cf. Pretlove et al. (2005)). If it is indeed the case that the temperatures experienced during the micro-climate of the bed exhibits less variation in the population than everyday temperatures, then measures of range or standard deviation in experienced temperature would characterise cold exposure, more than heat exposure, in the UK. Put another way, a large standard deviation measured on the AX3 is more likely to reflect a participant experiencing cold than heat, with the exception of frequent use of saunas or other unusually warm practices. It is this observation that suggests the standard deviation may be a possible characterisation of cold exposure for climates such as the UK. Moreover, such a measure does not suffer from the problem of arbitrariness that affects the first decile metric. The next chapter explores the metric t_{sd}, named here the *thermal variety*, to determine the extent to which this assertion is supported. However, before moving to the next chapter, the mean experienced temperature is considered.

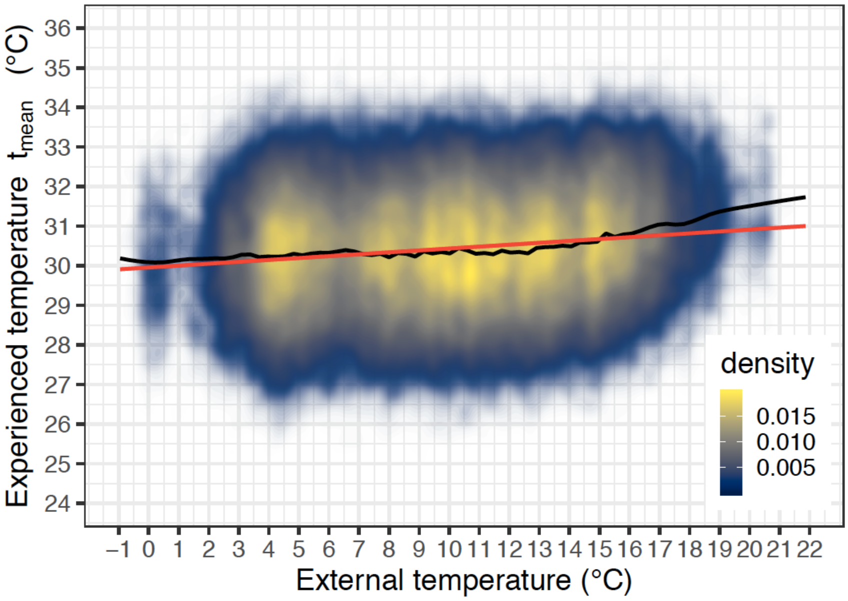


(a) The density distribution of all metrics: t^m_{min}, t^m_{10}, t^m_{25} and t^m_{μ}.

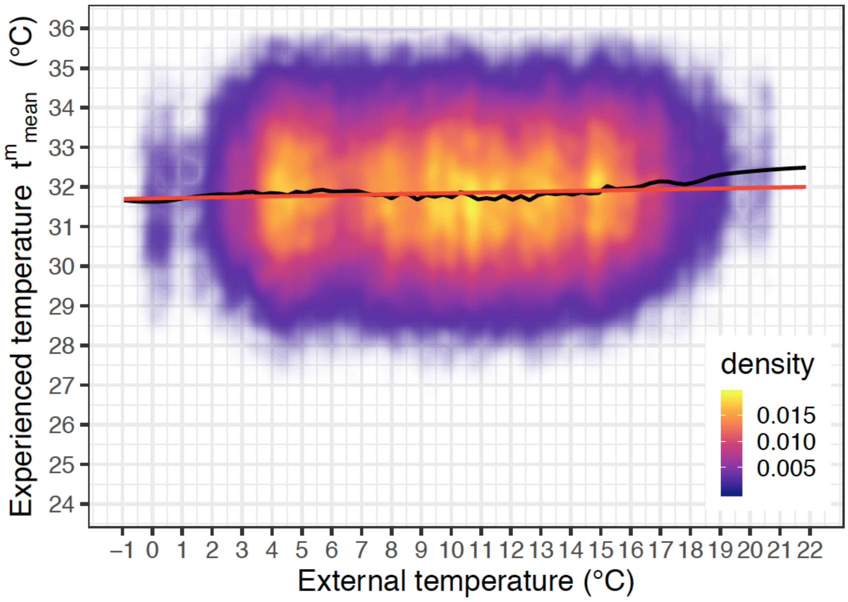


(b) The density distribution of all metrics: t_{min}, t_{10}, t_{25} and t_{μ}.

Figure 6.11: The density distributions of all metrics. The values of the metrics for which activity above the median level is excluded (superscript m) are warmer than those which are not filtered in such a manner.



(a) The relationship between external temperature and t_μ. The least squares fit of the data has gradient 0.05[0.05 – 0.05], $p < 2 \times 10^{-16}$).



(b) The relationship between external temperature and t^m_μ. The least squares fit of the data has gradient 0.01 [0.01 – 0.02], $p < 2 \times 10^{-16}$). The relationship is essentially flat.

Figure 6.12: The relationship between external temperature and t_μ (sub-figure (a)) and t^m_μ (sub-figure (b)). For both plots, the red line denotes the least squares fit, the black line is the LOESS regression as described in chapter 6.

6.6.1 Mean experienced temperature

The decision was made early on that the mean experienced temperature (t_μ) would not be a good a metric for summarising a time-series of each participant to capture cold exposure.

The following section addresses whether or not this was a good assumption. In the appendices, tables B.1, B.2 and B.3 lay out the results for the regression of t_μ and t_μ^m against the sociodemographic, building and health factors used in the study. Figure 6.13 gives the standardised effect size for the significant results for both t_μ and t_μ^m.

The first thing that is evident is that there are far fewer significant results, compared to the other metrics considered so far. In order to analyse the results clearly, first the variables which show similar significance for the two metrics are described. The intercept, which is equal to the overall average, for t_μ is 30.31 [30.23 – 30.39]°C, and for t_μ^m it is 32.02 [31.94 – 32.09]°C. This difference of 1.71°C reflects the average difference in mean experienced temperature between times when the participant was sedentary versus all other times.

For external temperature, there is a small but significant relationship for both metrics. For t_μ it is 0.05 [0.05 – 0.05]°C, and t_μ^m it is 0.01 [0.01 – 0.02]°C. These relationships are shown in the gradient of figures 6.12a and 6.12b. For both metrics there is a single significant difference as a function of age for those aged 70 – 79 versus those aged 40-49. For t_μ^m the difference is 0.11 [0.05 – 0.17]°C, and t_μ 0.13 [0.08 – 0.19]°C. The similarity of these readings is likely a result of the oldest people in the sample spending least time outdoors, so that the experienced temperature is approximately the same whether or not highest activity is removed.

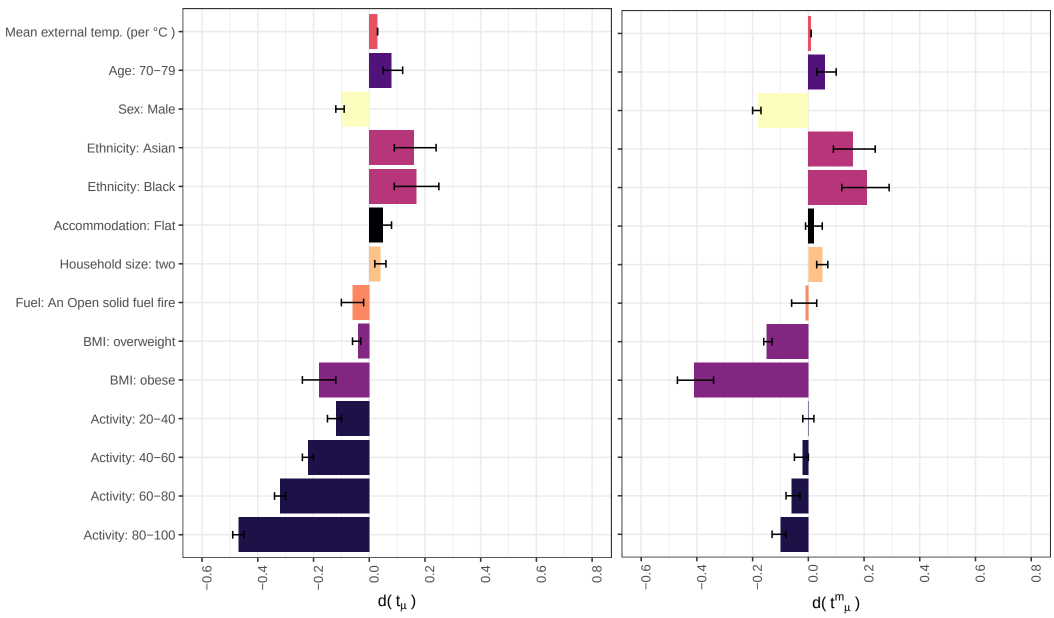


Figure 6.13: Standardised effect size estimates for t_μ and t_μ^m, for significant subcategories only. N=77,762. In order to calculate this, the estimates for t_μ^m were divided by the $sd(t_\mu^m) = 1.68$°C, and for t_μ by $sd(t_\mu) = 1.65$°C.

Males' mean experienced temperatures are significantly colder than females. For t_μ^m the difference is -0.31 [-0.33 – -0.28]°C and t_μ it is -0.17 [-0.19 – -0.14]°C. This is one of the few instances for which the effect size for a median screened metric is larger than the data set as a whole. This may be evidence of different thermal comfort preferences being evident. One might expect this to be especially the case for participants who live alone, since they establish their thermal environment without consideration of other occupants. However, restricting the analysis to only those who live alone (n=12,887) does not modify the results dramatically; for t_μ^m the difference between females and males is -0.27 [-0.33 – -0.21]°C, which is in line with the estimate based on the whole sample, and t_μ it is -0.12 [-0.18 – -0.06]°C.

As a function of household size itself, only the category of two person occupancy is significantly different from single occupancy. For t_μ, those living with one other person are 0.07 [0.03 – 0.10]°C warmer than those living alone. For t_μ^m the difference is almost identical at 0.08 [0.05 – 0.12]°C. It is difficult to interpret this result without more information on the physical parameters of the dwellings, but it may point to longer heating periods (as suggested above for the lower experienced temperature metrics).

There are two significant differences as a function of ethnic background. For those who identify their ethnic background as Black, t_μ^m is 0.34 [0.21 – 0.48]°C warmer than those who identify as White. The difference for t_μ is 0.28 [0.14 – 0.41]°C. A significant difference between those who identify their ethnic background as Asian, versus White, is also found. For t_μ^m it is 0.28 [0.15 – 0.41]°C and t_μ it is 0.27 [0.14 – 0.39]. This finding was not anticipated in either the literature review on wrist temperature or on internal temperatures. As with the previous results for ethnicity, it is difficult to interpret these findings but they may warrant further investigation.

There is evidence that mean experienced temperature is correlated with BMI. For t_μ^m, obese participants are 0.68 [0.58 – 0.79]°C colder than those who have a normal BMI. In terms of effect size this difference is 0.4, which is between small and moderate. For t_μ the effect is smaller at 0.30 [0.20 – 0.40]°C. This difference is also evident for overweight participants (t_μ^m=0.25 [0.22 – 0.27]°C colder than normal BMI, t_μ=0.07 [0.05 – 0.07]°C colder). Again, this is an example of where the effect size is greater for sedentary periods. There are several plausible factors that might influence experienced temperature measurements. Studies discussed in the literature review found that obesity was associated with lower wrist temperature with reduced amplitude of variation than those of normal weight (Corbalán-Tutau et al., 2011, 2015; Harfmann et al., 2017). The lower temperatures found in this study may be partially explained by these differences. Furthermore, a study of office based thermal comfort by Rupp et al. (2018) found that obese participants prefer cooler conditions. At least two possibilities arise from this finding, either obese participants choose cooler conditions or they wear systemically fewer clothes than participants of normal BMI – either of these possibilities is consistent with the finding that the experienced temperature of obese participants is significantly lower than those of participants with normal BMI.

Clear effects are evident as a function of activity level. Those in the highest quintile have a t_μ 0.78 [0.72 – 0.84]°C lower than those in lowest activity quintile. Intermediate quintiles show a similar trend, with a decrease of approximately 0.16°C per quintile. This result is consistent with the observation that most physical activity takes place outdoors and/or with less clothing than at sedentary times. The results for t_μ^m is more complex, a significant effect only begins at the fourth quintile, with a difference of -0.09 [-0.13 – -0.05]°C compared to the first quintile. The fifth quintile is -0.17 [-0.21 – -0.13]°C. It is important not to discount the potential impact of length of sleeping period. Those with higher recorded activity levels may well be awake for longer than those with lower values – since the micro-climate of the bed is warmer than most ambient room temperature in the UK this would result in lower t_μ and t_μ^m readings for most active people. This point will be returned to in section 10.2.

There are two results that are significantly different for t_μ^m but not for t_μ, both in the tenure type variable. Compared to those who own their home outright, those with a mortgage are 0.06 [-0.09 – -0.03]°C colder, and those who rent privately are 0.14 [-0.23 – -0.05]°C colder. There is some evidence that those living in fuel poverty are more likely to live in privately rented accommodation, which could contribute to heat rationing and reduced indoor temperatures (DECC, 2015).

With the addition of the variables *financial situation satisfaction*, *health satisfaction* and *heating type* there was only one significant difference for either metric, for t_μ^m, those who were very unhappy with their financial situation were 0.19 [0.05 – 0.34]°C warmer. Given that none of the other subcategories in this variable showed a significant result it is difficult to interpret.

6.7 Summary

This chapter has reported the results pertaining to research question 1, which asked whether experienced temperature varies with sociodemographic, health and building factors. The lower metrics of experienced temperature show significant variation with sociodemographic and building factors. Less variation was shown using mean experienced temperature. The following chapter considers metrics of thermal variety and their relationship with the same factors considered in this chapter. These results, together with those presented in chapter 8, will be discussed as a whole in chapter 9 and considered in relation to the hypotheses set out in chapter 3 and the literature more widely.

Chapter 7

Results 3: Thermal variety

In a country where the winters are so cold as in Great Britain, fuel is, during that season, in the strictest sense of the word, a necessary of life

Adam Smith – Wealth of Nations (1776)

This chapter introduces and explores metrics which capture the diversity of temperatures experienced by participants. It focuses principally on the standard deviation of the experienced temperature, named the thermal variety, and denoted t_{sd}, along with the analogue to t_{10}^{m} denoted t_{sd}^{m}. A brief background justification of why the thermal variety is an independent and useful extension of the metrics considered so far is also given. The results of using t_{sd} as the outcome variable are then examined, as was done in chapter 6 for the lower metrics of experienced temperature. Two alternative metrics of diversity are also considered, namely the range and interquartile range. Finally, the metrics are examined together, to determine which best characterises cold exposure and is most useful for the research aims of this study.

7.1 Thermal variety

The variation in ambient temperature enters the epidemiological literature principally through the concept of Diurnal Temperature Range (DTR) (see section 2.1.2). The review by Cheng et al. (2014), for example, found that DTR has been associated with negative health outcomes, especially for cardiovascular and respiratory conditions. However, the impacts of DTR may be confounded by socioeconomic status and season. The equivalent of DTR for this study is a measure of variation of experienced temperature.

The thermal variety was computed using the AX3 time-series data. As with the other metrics analysed thus far, the first and last days of the time-series recorded by the AX3 were removed to minimise the impact of end effects of the periods when the AX3 was unworn. Before considering whether or not the t_{sd} and t_{sd}^{m} metrics are a useful means of establishing the relationship between experienced temperature and health, the question of the extent to which they constitute metrics in their own right, independent of t_{10} and the other metrics considered thus far.

7.1.1 An independent measure?

The extent to which thermal variety and the first decile of experienced temperature are independent of each other is shown in figure 7.1. Figure 7.1a shows a negatively correlated linear relationship between t_{10} and t_{sd}, with a large degree of scatter. The equation for the linear relationship is $t_{sd} = -0.21t_{10} + 6.84$, $R^2 = 0.40$. The value of $R^2 = 0.40$ suggests the two metrics are not wholly independent. The LOESS fit is also shown, which shows some deviation from linearity at extremal values of t_{10}, but the values at high t_{10} are sparse. There is a clear colour gradient evident, suggesting that high thermal variety t_{sd} and lower experienced temperature t_{10} is associated with high activity levels.

Figure 7.1b again shows a negatively correlated linear relationship, this time between t_{10}^m and t_{sd}^m again with a large degree of scatter. The equation for the linear relationship is $t_{sd}^m = -0.15t_{10}^m + 6.84$, $R^2 = 0.26$. The LOESS fit is also shown, which shows some deviation from linearity at extremal values of t_{sd}^m. The value of $R^2 = 0.26$ suggests the two metrics are more independent than t_{sd} and t_{10}, but that including both in a single regression would likely cause issues of collinearity. The absence of a clear pattern in the colouring of the data points by activity levels suggests that removing the data points within a particular time-series which correspond to high activity is an adequate method of screening times of high activity from the analysis.

For a normally distributed variable x the following relation holds for the n^{th} percentile,

$$x_{sd} = \frac{x_n}{z_n} - \frac{x_\mu}{z_n} \tag{7.1}$$

where z is the z-score for a particular percentile n ($z = -1.282$ for $n = 10$). Comparing the form of this equation with those above, using the computed values of t_μ^m and t_μ respectively, reveals the following differences in coefficients.

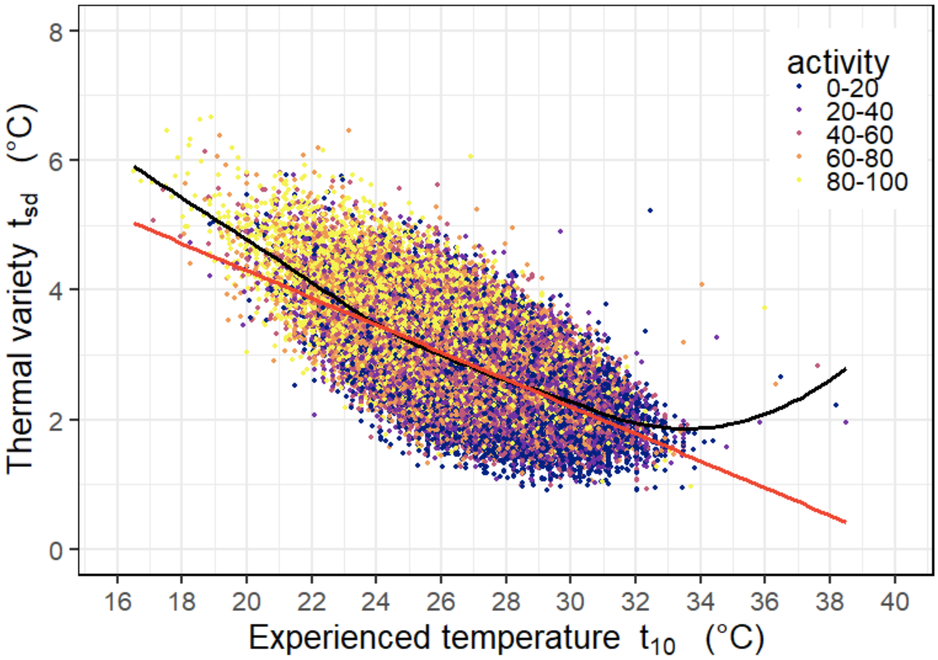
$$\text{observed :} \quad t_{sd}^m = -0.15t_{10}^m + 6.84 \tag{7.2}$$

$$\text{normal :} \quad t_{sd}^m = -0.78t_{10}^m + 24.86 \tag{7.3}$$

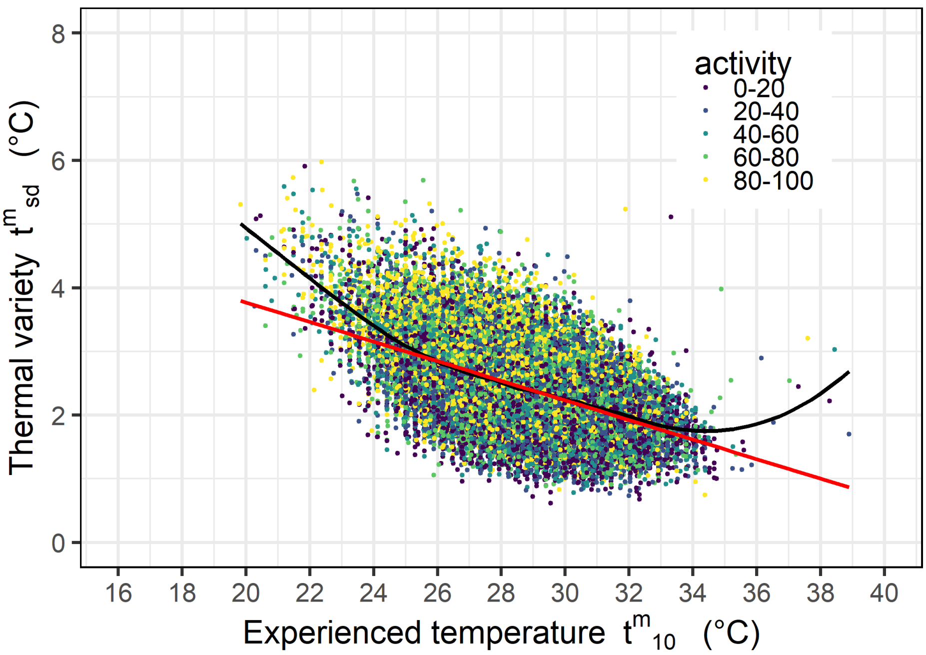
$$\text{observed :} \quad t_{sd} = -0.21t_{10} + 8.49 \tag{7.4}$$

$$\text{normal :} \quad t_{sd} = -0.78t_{10} + 23.74 \tag{7.5}$$

These comparisons suggest that the data for t_{sd}^m and t_{10}^m are further from normally



(a) $t_{sd} = -0.21t_{10} + 8.49$, $R^2 = 0.40$. ($p < 2 \times 10^{-16}$)



(b) $t^m_{sd} = -0.15t^m_{10} + 6.84$, $R^2 = 0.26$. ($p < 2 \times 10^{-16}$),

Figure 7.1: Comparing the relationships between the thermal variety and experienced temperature for the unscreened data, and those which have been screened to exclude activity levels above the median.

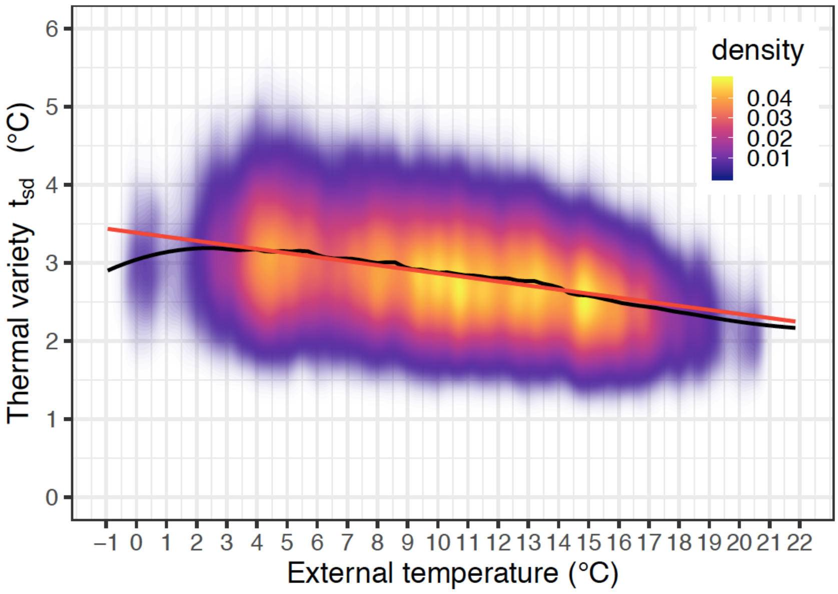


Figure 7.2: The relationship between external temperature and t_{sd}. The red line is the least squares fit of the data (gradient -0.05 [-0.05 – -0.05], $p < 2 \times 10^{-16}$). The black line is the LOESS regression as described in chapter 6. Since over 77,762 data points are plotted, the data are represented as density cloud, as given in the key

distributed than the data for t_{sd} and t_{10}. Since the former are subset of the latter, produced by removing periods of time associated with high activity, this stands to reason. The next section gives the results of including t_{sd} and t^m_{sd} as the outcome variable of a multiple linear regression model, in the same manner as used above for the alternative metric, but with more detail. The structure of the results presented here is similar to that of the previous results chapter. First, the variation of thermal variety with external temperature is given in 7.2. This shows a clear effect, with a higher thermal variety evident at the colder times of the year than the warmer. This accords with the observation that the difference between external and internal temperatures is greatest during the winter months in Britain, and so participants experience more thermal variety during these months.

7.1.2 Multiple linear regression

This section follows the approach given in chapter 6. The results of the multiple regression model with t_{sd} and t^m_{sd} used as outcome variables are given in tables 7.1 and 7.2, with the results of the additional explanatory variables given in 7.3.

In line with figure 7.2, the thermal variety decreases with external temperature. The estimates for t_{sd} and t^m_{sd} are similar, at -0.05°C and -0.04°C respectively. It is notable that t^m_{sd} is higher in winter. This suggests increasing thermal variety during sedentary periods as the average external temperature drops. The relative effect size for both metrics is given below in section 7.1.3.

As age increases, the thermal variety decreases monotonically, again with similar magnitude for t_{sd} and t_{sd}^m. The explanation for this is likely the same as that given above for the minimum temperature variables – as a function of age, the differences between the median activity filtered data and the unfiltered data are small.

There is a small but significant difference in thermal variety as a function of sex. Men have a lower thermal variety than women. The difference is greater for t_{sd}^m in both absolute and relative terms (as an effect size, see table B.8). It is difficult to speculate what the basis for this observed sex difference might be.

As with other metrics, some differences are observed as a function of ethnicity. For t_{sd} only those whose ethnic background is Black are significantly different from White participants. The thermal variety of Black participants is 0.09°C higher than White. This finding is reflected in t_{sd}^m which finds the same magnitude of difference. Two further differences are recorded; participants who identify their ethnic background as Chinese have higher thermal variety during likely sedentary periods of 0.20°C, and those who describe their background as "other ethnic background" also show increased thermal variety under this metric. As with the differences highlighted in the previous section, it is very difficult to interpret these results without further information.

Notably, there were no significant differences as a function of household income. This contrasts the findings for t_{min} and t_{min}^m, which found that higher income participants had lower minimum temperatures. This will be discussed further in section 7.1.4 below.

Those living in flats show lower thermal variety, under both metrics, than those living in houses or bungalows. This may relate to the building properties of flats, for which the internal temperatures likely vary less than for other accommodation types due to the proximity of other flats which would moderate large changes in temperature. It could also be that flats are much more likely to be located in urban and semi urban areas, which experience lower swings in temperature (although higher absolute temperatures) due to the urban heat island effect.

A similar argument might account for the observed differences by tenure type; those renting from the Local Authority have lower variety than those who own their home outright. The build quality of Local Authority properties is also generally higher than privately owned homes. The differences observed for those who have mortgages are small and difficult to account for. There is a small but clear effect as a function of household size. Participants living in homes with two, three or four or more inhabitants show significantly lower thermal variety than those living alone.

Those who are unable to work due to sickness or disability have slightly reduced t_{sd}^m relative to those who work. Those doing voluntary work have both reduced t_{sd} and t_{sd}^m, and those for whom none of the available categories apply have slightly increased thermal variety for both metrics. It is not clear why the thermal variety of those doing voluntary work would

Predictor (relative subcategory, N)	Sub-category (N)	t_{sd}	t^m_{sd}
Intercept	-	3.46 [3.43 – 3.48] ***	3.10 [3.08 – 3.13] ***
External temperature	-	-0.05 [-0.06 – -0.05] ***	-0.04 [-0.04 – -0.04] ***
Age (40-49, 6075)	50-59 (21320)	-0.06 [-0.08 – -0.04] ***	-0.06 [-0.08 – -0.05] ***
	60-69 (35407)	-0.10 [-0.12 – -0.08] ***	-0.12 [-0.14 – -0.10] ***
	70-79 (14960)	-0.16 [-0.18 – -0.14] ***	-0.15 [-0.17 – -0.13] ***
Sex (Female, 43770)	Male (33992)	-0.05 [-0.06 – -0.04] ***	-0.09 [-0.10 – -0.09] ***
Ethnic background (White, 75365)	Mixed (398)	0.07 [0.01 – 0.13]	0.07 [0.02 – 0.12] *
	Asian (654)	-0.01 [-0.05 – 0.04]	0.03 [-0.01 – 0.07]
	Black (582)	0.09 [0.04 – 0.14] **	0.09 [0.05 – 0.13] **
	Chinese (157)	0.11 [0.02 – 0.21]	0.20 [0.12 – 0.28] **
	Other ethnic group (395)	0.03 [-0.03 – 0.09]	0.09 [0.04 – 0.14] **
	Do not know (20)	0.08 [-0.18 – 0.35]	0.11 [-0.11 – 0.34]
	Prefer not to answer (191)	-0.05 [-0.13 – 0.04]	-0.01 [-0.09 – 0.06]
Household Income £ (Less than 18,000, 10592)	18,000 to 30,999, (17779)	-0.02 [-0.04 – -0.01] *	-0.02 [-0.03 – -0.01] *
	31,000 to 51,999 (20016)	-0.01 [-0.03 – 0.00]	-0.01 [-0.02 – 0.01]
	52,000 to 100,000 (17021)	-0.01 [-0.03 – 0.01]	-0.01 [-0.02 – 0.01]
	Greater than 100,000 (4850)	-0.02 [-0.04 – 0.01]	-0.01 [-0.03 – 0.01]
	Prefer not to answer (5475)	-0.01 [-0.03 – 0.01]	-0.02 [-0.04 – -0.00]
	Do not know (2029)	-0.07 [-0.10 – -0.04] **	-0.06 [-0.08 – -0.03] **
Accommodation type (House/bungalow, 71554)	Flat (6058)	-0.07 [-0.09 – -0.05] ***	-0.05 [-0.06 – -0.03] ***
	Temporary (54)	0.02 [-0.14 – 0.18]	-0.03 [-0.17 – 0.11]
	None of above (83)	-0.05 [-0.18 – 0.08]	-0.05 [-0.16 – 0.06]
	Prefer not to answer (13)	-0.17 [-0.51 – 0.17]	-0.22 [-0.51 – 0.07]
Tenure type (Own outright, 44537)	Mortgage (28498)	-0.05 [-0.07 – -0.04] ***	-0.03 [-0.04 – -0.02] ***
	Rent Local Authority (2096)	-0.16 [-0.18 – -0.13] ***	-0.11 [-0.14 – -0.09] ***
	Rent private (1497)	-0.04 [-0.07 – -0.01]	0.00 [-0.03 – 0.03]
	Shared (174)	-0.07 [-0.16 – 0.02]	-0.06 [-0.14 – 0.01]
	Rent free (469)	-0.09 [-0.15 – -0.04]	-0.02 [-0.07 – 0.02]
	None of above (276)	-0.07 [-0.14 – 0.00]	-0.02 [-0.08 – 0.04]
	Prefer not to answer (215)	-0.01 [-0.09 – 0.08]	0.01 [-0.07 – 0.08]
Household size (single occupant, 12854)	Two (37905)	-0.04 [-0.05 – -0.02] **	-0.07 [-0.08 – -0.06] ***
	Three (12141)	-0.05 [-0.06 – -0.03] ***	-0.06 [-0.07 – -0.05] ***
	Four or more (14862)	-0.03 [-0.05 – -0.01] ***	-0.04 [-0.06 – -0.03] ***

Table 7.1: The results of the regression model of the associations between t_{sd} and explanatory variables, and t^m_{sd} and explanatory variables, as described in the text. Significance levels: * $p < 0.01$, ** $p < 0.001$, *** $p < 1 \times 10^{-9}$. N=77,762

be significantly different from other categories. The category size is large enough (n=3,759) that this result would seem statistically valid.

As with the findings for t^m_{min} and t_{min}, the presence of an open solid fuel fire is associated with increased thermal variety. Unlike the findings for the minimum temperature metrics, both thermal variety metrics, t^m_{sd} and t_{sd}, are significantly higher when an open fuel solid fire is present in all such subcategories.

The results also suggest a coherent picture for BMI in terms of thermal variety. All subcategories are significantly different from normal BMI. Those who are underweight have higher thermal variety. Those who are overweight and obese have lower thermal variety. In this sense, BMI is monotonically inversely related to both t_{sd} and t^m_{sd}. These findings accord with the papers due to Nissilae et al. (1996) and Harfmann et al. (2017). This suggests a degree of decreased variety is likely due to reduced wrist temperature variation. However, the

Predictor (relative subcategory, N)	Sub-category (N)	t_{sd}	t^{m}_{sd}
Employment status (In paid employment or self-employed, 39797)	Retired (27472)	0.03 [-0.03 – 0.09]	0.02 [-0.03 – 0.08]
	Looking after home/family (3235)	0.03 [-0.09 – 0.15]	-0.01 [-0.12 – 0.09]
	Unable to work, sickness/disability (1411)	0.01 [0.00 – 0.02]	-0.02 [-0.03 – -0.01] **
	Unemployed (901)	0.02 [-0.00 – 0.04]	0.02 [-0.00 – 0.04]
	Doing unpaid or voluntary work (3759)	-0.10 [-0.13 – -0.06] **	-0.05 [-0.08 – -0.03] **
	Full/ part-time student (738)	-0.02 [-0.06 – 0.02]	0.00 [-0.04 – 0.03]
	None of the above (350)	0.04 [0.01 – 0.06] **	0.03 [0.01 – 0.05] *
	Prefer not to answer (99)	0.04 [-0.00 – 0.08]	0.04 [0.00 – 0.08]
Fuel type (Gas hob or gas cooker, 28957)	Gas fire (6379)	0.01 [-0.00 – 0.03]	0.01 [-0.01 – 0.02]
	Open solid fuel fire (2335)	0.12 [0.09 – 0.14] ***	0.07 [0.05 – 0.09] ***
	Gas hob & Gas fire (20188)	0.01 [-0.00 – 0.02]	0.01 [-0.00 – 0.01]
	Gas hob & Open fire (4481)	0.09 [0.07 – 0.11] ***	0.06 [0.05 – 0.08] ***
	Gas fire & Open fire (195)	0.21 [0.12 – 0.29] **	0.19 [0.12 – 0.27] **
	Gas hob & Gas fire & Open fire (956)	0.08 [0.04 – 0.12] **	0.09 [0.06 – 0.12] **
	None of the above (14221)	-0.01 [-0.02 – 0.00]	-0.01 [-0.02 – -0.00]
	Prefer not to answer (37)	-0.21 [-0.41 – -0.01]	-0.20 [-0.37 – -0.03]
	Do not know (13)	-0.18 [-0.50 – 0.15]	-0.15 [-0.43 – 0.14]
Body Mass Index (Normal, 30562)	Underweight (477)	0.11 [0.06 – 0.17] **	0.11 [0.07 – 0.16] **
	Overweight (45722)	-0.18 [-0.19 – -0.18] ***	-0.14 [-0.15 – -0.13] ***
	Obese (1001)	-0.37 [-0.41 – -0.34] ***	-0.24 [-0.27 – -0.21] ***
Activity level quintile (1st quintile, 15463)	2nd quintile (15567)	0.14 [0.13 – 0.16] ***	0.04 [0.03 – 0.05] ***
	3rd quintile (15567)	0.24 [0.22 – 0.25] ***	0.08 [0.07 – 0.09] ***
	4th quintile (15578)	0.33 [0.31 – 0.34] ***	0.11 [0.10 – 0.12] ***
	5th quintile (15587)	0.50 [0.49 – 0.51] ***	0.19 [0.18 – 0.20] ***

Table 7.2: The continuation of table 7.1. The results of the regression model of the associations between t_{sd} and explanatory variables, and t^m_{sd} and explanatory variables, as described in the text. Significance levels: * $p < 0.01$, ** $p < 0.001$, *** $p < 1 \times 10^{-9}$. N=77,762

magnitude of the differences observed by the present study cannot be accounted for by wrist temperature changes alone. This suggests that both the participants' thermal environment and their wrist temperature variation are associated in a coherent way with BMI level.

In a manner that agrees with the previous metrics, activity level shows a significant positive association with both t_{sd} and t^m_{sd}. Again, the simple explanation for this is that more active people have access to a wider variety of thermal environments.

The single significant difference as a function of *financial situation satisfaction* does not lend itself to coherent interpretation. On the other hand, *health satisfaction* shows a much clearer trend. Compared to participants who are extremely satisfied with their health, those who are increasingly unhappy show diminished thermal variety, for both metrics.

Finally, there is evidence that those who have oil central heating systems installed in their home have greater thermal variety, again for both metrics. The explanation for this likely lies in the observation that oil central heating systems are more common in rural areas, in older, larger, less well insulated homes.

7.1.2.1 Statistical checks

As with the previous regression models, it is important to check that the residuals are normally distributed. These are given as histograms in figure 7.3.

Figure 7.2 shows the relationship between mean external temperature and t_{sd}. While there is slight deviation from linearity at very low average temperature, it is likely insufficient

Predictor (relative subcategory, N)	Sub-category (N)	t_{sd}	t^m_{sd}
Financial situation satisfaction (Extremely happy, 3808)	Very happy (14498)	0.01 [-0.01 – 0.04]	0.02 [-0.00 – 0.04]
	Moderately happy (15732)	0.01 [-0.01 – 0.03]	0.02 [0.00 – 0.04]
	Moderately unhappy (2473)	-0.02 [-0.05 – 0.02]	0.01 [-0.02 – 0.04]
	Very unhappy (737)	-0.07 [-0.12 – -0.02] *	-0.03 [-0.07 – 0.02]
	Extremely unhappy (369)	-0.06 [-0.13 – 0.00]	-0.03 [-0.09 – 0.03]
	Prefer not to answer (57)	-0.09 [-0.25 – 0.07]	-0.02 [-0.16 – 0.11]
	Do not know (56)	-0.12 [-0.28 – 0.03]	-0.09 [-0.23 – 0.05]
Health satisfaction (Extremely happy, 2230)	Very happy (13771)	-0.04 [-0.07 – -0.01] *	-0.02 [-0.04 – 0.00]
	Moderately happy (17767)	-0.10 [-0.12 – -0.07] ***	-0.05 [-0.07 – -0.03] ***
	Moderately unhappy (2955)	-0.15 [-0.19 – -0.12] ***	-0.09 [-0.12 – -0.06] ***
	Very unhappy (661)	-0.16 [-0.21 – -0.11] ***	-0.08 [-0.13 – -0.03] **
	Extremely unhappy (249)	-0.15 [-0.23 – -0.07] **	-0.07 [-0.14 – -0.00]
	Prefer not to answer (10)	0.05 [-0.32 – 0.42]	-0.05 [-0.37 – 0.27]
	Do not know (87)	-0.04 [-0.17 – 0.08]	-0.03 [-0.14 – 0.09]
Heating type (Gas central heating, 34999)	Electric storage heaters (798)	-0.01 [-0.05 – 0.03]	-0.02 [-0.06 – 0.02]
	Oil (kerosene) central heating (979)	0.09 [0.05 – 0.13] **	0.07 [0.03 – 0.10] **
	Portable gas or paraffin heaters (10)	0.17 [-0.20 – 0.54]	0.04 [-0.28 – 0.36]
	Solid fuel central heating (128)	0.09 [-0.01 – 0.20]	-0.01 [-0.10 – 0.08]
	Open fire without central heating (109)	-0.02 [-0.14 – 0.09]	-0.01 [-0.11 – 0.09]
	Three heating types (5)	-0.17 [-0.69 – 0.35]	-0.05 [-0.50 – 0.40]
	None of the above (676)	-0.01 [-0.05 – 0.04]	0.00 [-0.04 – 0.04]
	Prefer not to answer (15)	-0.19 [-0.53 – 0.16]	-0.22 [-0.52 – 0.07]
	Do not know (11)	-0.17 [-0.52 – 0.19]	-0.19 [-0.49 – 0.12]

Table 7.3: The results of adding the *health satisfaction*, *financial situation satisfaction* and heating type variables to the regressions models reported in tables 7.1 and 7.2. These explanatory variables were only available for a smaller subset of participants. While all variables were included, only *Financial situation satisfaction* and *Heating type* are shown. Again, the other variables agreed with the previous regression to within the confidence intervals. Significance levels: * $p < 0.01$, ** $p < 0.001$, *** $p < 1 \times 10^{-9}$.. N=37,730

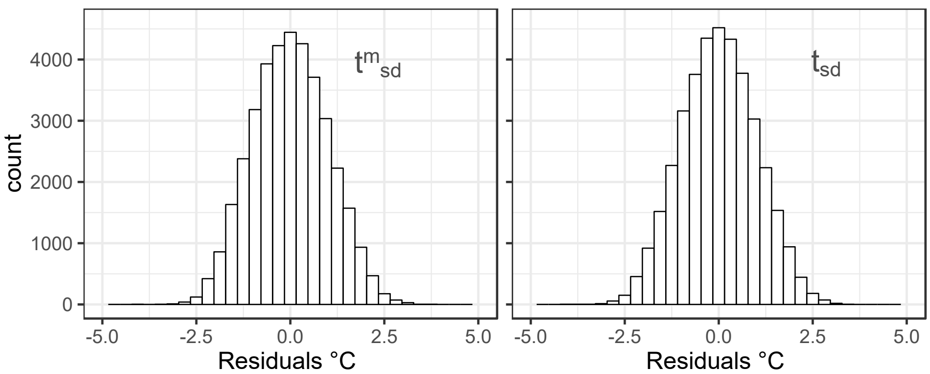


Figure 7.3: The residuals for the above fits of t^m_{sd} and t_{sd}. They are both normally distributed. The residuals for t^m_{sd} have a slight positive skew which is probably insufficient to cause problems.

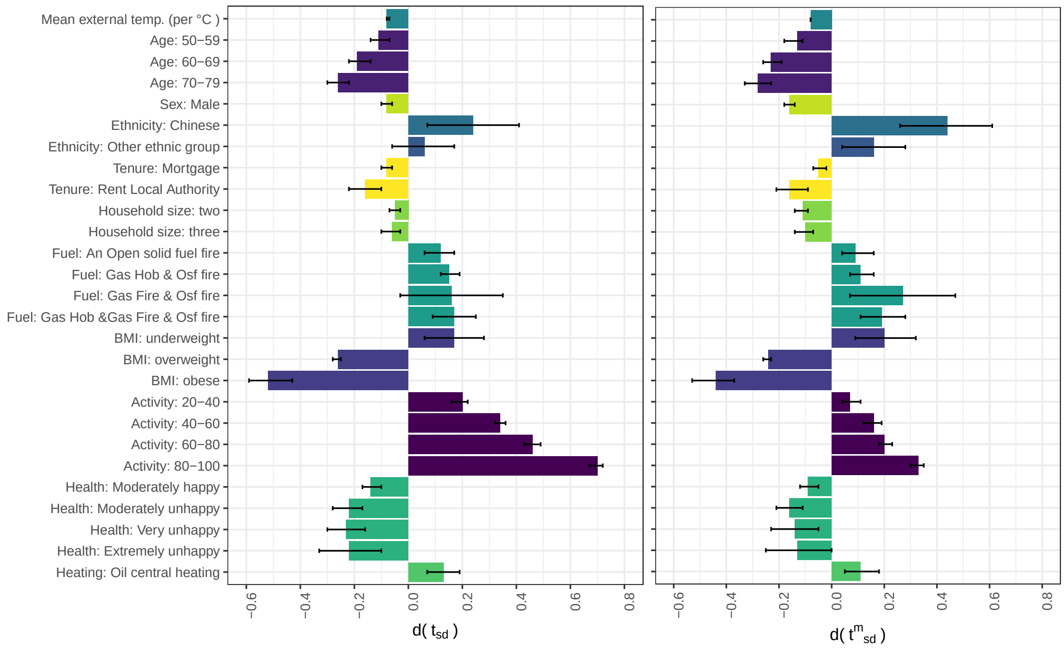


Figure 7.4: Effect size estimates for t_{sd}^{m} and t_{sd}.

to affect the model. It is therefore reasonable to conclude that the models are valid from a statistical standpoint. The next section considers the effect size of the observed statistically significant differences using t_{sd} and t_{sd}^{m} as outcome variables.

7.1.3 Effect size

As with the results given in chapter 6, it is instructive to examine the effect size of the associations revealed by the multiple regression model. These are plotted in figure 7.4. Over a total range of a 12°C change between summer and winter external temperatures, t_{sd} is 0.6°C lower in summer than winter. This equates to an effect size of 0.87, which is large under the rule of thumb given by Sawilowsky (2009).

Overall, aside from average external temperature, the largest effect sizes are observed for the activity and BMI variables, followed by age and *health satisfaction*. Typically the effect sizes for t_{sd} are larger than for t_{sd}^{m}. This likely is the result of the restriction of scope of the t_{sd}^{m} variable, since it only makes use of half the data of the t_{sd} variable.

7.1.4 Alternative measures of variety

There are other possible measures of variety. Two which will be briefly discussed here are the range and the interquartile range. These are defined as the difference between the minimum and maximum values, and the difference between the first and third quartiles, respectively. These are denoted t_{range} and t_{iqr}. All significant differences using these metrics as the outcome variable in the multiple linear regression are given in appendix tables B.4, B.5 and B.6. Examining these variables sheds light on the reason why t_{min} was found to be associated with income level but t_{sd} was not. Since table B.1 shows that t_{max} increases with increasing income, it follows that t_{range} also shows this relationship – indeed, increased income is associated with *increasing* thermal range. However, examining the t_{iqr}, the opposite effect is observed. Increasing income is associated with *decreasing* interquartile range. This suggests that while the extremal values expand with income level, the central 50% of the readings get closer together. It is for this reason that no effect is observed as a function of t_{sd}, the results for range and interquartile range effectively cancel each other out. A similar result is observed as a function of sex – the range of experienced temperatures for males is greater than females, but the interquartile range is reduced. However, there is also a small reduction in standard deviation for males versus females, suggesting the effects do not cancel each other out to the same extent as the income variable. No other variables show this pattern, and the results for t_{range} and t_{iqr} generally agree with the findings for t_{sd}. For median filtered metrics, a similar result holds, but as discussed earlier, the effect size tends to be smaller. The next section addresses the binomial regression model, as was discussed in chapter 7, and the impact of using measures of variety in it.

7.2 Robustness tests

The following section examines the robustness of the results presented so far and considers some potential sources of error.

7.2.0.1 Weekends

As discussed in the literature review, a study by Huebner et al. (2013) found a significant difference of 0.16°C between weekday and weekend temperatures of households. The AX3 wristbands were distributed at random to participants. However, this study only uses 5 of the total 7 days due to the problems at the start and end of the wear period. It is plausible that the weekend might be systematically under or over sampled as a result. A total of 18,223 participants had monitoring periods which did not include any weekend days. Therefore, it might be expected that monitoring periods which included weekends might be significantly warmer to those which only included week days. This was tested in the data for all the experienced temperature and thermal variety metrics by re-running the regression models with an extra variable which counted the number of weekend days in the study period. The result of these tests was that no significant differences were found.

7.2.0.2 External temperature

Since the experienced temperature depends most heavily on local average external temperature, a further regression of the binomial regression model was produced for which the thermal variety metrics were corrected for external temperature, to minimise the risk of multiple co-linearity effecting the model. The inclusion of these modified metrics had no

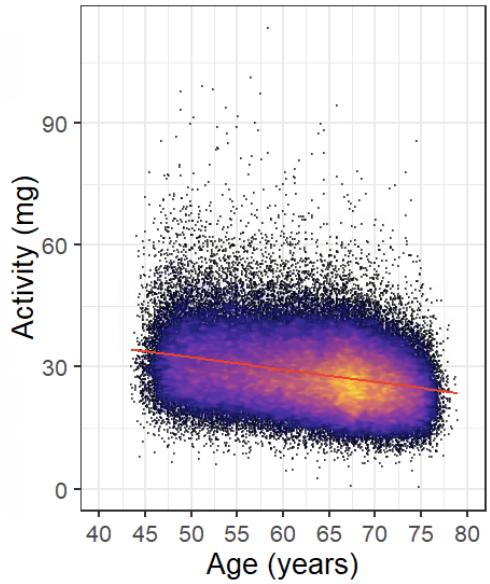


Figure 7.5: The relationship between age and mean activity level. Simple least-squares regression shown in red. Points represented as density cloud.

effect on the results.

7.2.0.3 Participant exclusions

As part of the Pre–analysis plan (PAP), a number of participants were excluded out of concern that abnormally cold hands might bias the result. The decision was made that self-reporting of conditions would be the basis of these exclusions. However, the numbers of participants who self-reported anaemia was much lower (n=262) than the number who were diagnosed with such conditions (n> 3000). As a check of the impact of the decision, those who had been diagnosed with anaemia were excluded and the analysis re-run. The results were strikingly similar to those reported above, all estimates agreed to within confidence intervals given, and almost all did not even change the numerical value of the estimate.

A related aspect to this is a validation of the dataset that can be performed by relating it to the results of Doherty et al. (2017). For this study, the relationship between activity level and other factors was considered. For comparison, the simplest one to perform is the relationship with age. The general relationship is shown in figure 7.5. Doherty et al. (2017) use slightly different age factoring brackets to those used here, so the data from this study was refactored to compare it. The results of this comparison are shown in table 7.4. The results demonstrate that, despite the different sample size (due to the different exclusion criteria placed on the data for this study) they agree.

Study	Age	Female	Male
Doherty et al. (2017)	45-54	31.2 ± 8.7	31.1 ± 9.7
	55-64	29.1 ± 8.0	28.8 ± 8.8
	65-74	26.6 ± 7.1	25.6 ± 7.7
	75-79	23.9 ± 6.5	22.9 ± 6.8
Present study	45-54	33.4 ± 9.2	32.7 ± 7.5
	55-64	30.6 ± 7.9	30.0 ± 8.5
	65-74	28.1 ± 7.1	26.7 ± 7.7
	75-79	25.1 ± 6.6	23.8 ± 6.5

Table 7.4: The mean activity level $\pm$ standard deviation for the Doherty et al. (2017) (above) study compared to the present study (below). The very slight difference may be accounted for by the different sample sizes (n = 103,578 for Doherty et al. and n = 78,210 present study), as well as the different exclusion criteria and slightly different processing techniques.

7.3 Results overview

The following section reviews the results presented in the present chapter and the previous for the multiple linear regression models of the associations between the metrics of experienced temperature and thermal variety and the explanatory sociodemographic, housing and health factors. This is done in two ways, starting with a discussion of the R^2 value for each model, and then followed by an analysis of the number of significant variables for each model.

7.3.1 Variance explained

One method of comparing the different regression models is to calculate the partial R^2 for each variable in each regression. This gives an estimate for the percentage of total variance explained by each variable in the model. Table 7.5 gives the values of partial R^2 for each variable in each model, as well as R^2 for the model as a whole. The values of R^2 have been corrected for the number of variables in each model.

Overall, the total explained variance is largest for the models which have t_{sd} as the outcome variable. Models for t_{10} have a greater degree of explained variance than t_{10}^m. The addition of the extra variables of *activity*, *BMI* and *health satisfaction* increase the amount of variance explained.

Table 7.5 also shows that the degree of variance explained is greatest for the external temperature variable in every model (except t_μ, which shows the greatest value for activity level). This is likely due to the physical nature of external temperature – sociodemographic variables are typically less predictive than physical variables in models which predict physical quantities such as the experienced temperature. The external temperature was also measured at the same time as the experienced temperature. Partial R^2 is also greater for the models which use the unscreened temperature compared to the data screened for high activity levels (i.e. t_{10} compared to t_{10}^m). This may be because internal temperatures, where people are typically less active, are generally warmer and

Explanatory variable	t^m_{10} (PAP)	t_{10} (PAP)	t^m_{min}	t_{min}	t^m_{10}	t_{10}	t^m_{μ}	t_{μ}	t^m_{sd}	t_{sd}
External temperature °C	3.93	6.54	8.93	12.76	4.12	7.44	0.15	2.04	12.42	14.17
Age/year	0.28	0.71	-	-	-	-	-	-	-	
Age (decade)	-	-	0.70	0.80	0.14	0.20	0.03	0.05	0.42	0.36
Sex	0.04	0.00	0.19	0.80	0.07	0.06	0.75	0.24	0.73	0.19
Ethnicity	0.03	0.03	0.03	0.15	0.06	0.04	0.13	0.11	0.10	0.02
Household income	0.02	0.01	0.05	0.13	0.06	0.01	0.00	0.00	0.03	0.04
Tenure Type	0.02	0.12	0.00	0.02	0.00	0.02	0.01	0.00	0.10	0.15
Accommodation type	0.03	0.07	0.00	0.00	0.02	0.05	0.00	0.00	0.06	0.05
Household size	0.11	0.03	0.04	0.04	0.01	0.03	0.04	0.00	0.16	0.04
Employment status	0.03	0.13	0.05	0.03	0.02	0.03	0.01	0.02	0.08	0.02
Fuel Type	0.09	0.24	0.07	0.13	0.05	0.11	0.00	0.00	0.19	0.26
BMI	-	-	0.08	0.13	0.03	0.12	0.58	0.07	1.73	2.17
Activity decile	-	-	3.18	5.90	0.64	5.04	0.14	2.36	1.31	6.04
Health satisfaction	-	-	0.05	1.70	0.00	0.06	0.01	0.00	0.15	0.37
Financial situation satisfaction	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.04
Heating type	-	-	0.02	0.00	0.00	0.02	0.00	0.00	0.02	0.04
Total	4.88	8.08	14.50	21.02	5.54	13.56	2.21	4.84	18.61	24.26

Table 7.5: Partial R^2 expressed as a percentage for each variable and model. Each column gives the explanatory variables used in each model. . The total R^2 is given at the bottom and may not be a sum of individual partial R^2 values. All values are corrected for the number of explanatory variables in the model. As a test, the first column was run with age in continuous and discrete forms (see section 8.3. The impact was minimal, a maximum change of 0.01 for non-age variables, and 0.23% instead of 0.28% for the age variable.

show less variance than external ones.

The external temperature does not explain t^m_μ very well. The total explained variance for t^m_μ is only 2.21%, with the variable *sex* having the largest partial R^2. This finding supports the suggestion made in the PAP that the readings of t^m_μ might be dominated by the higher temperature readings from night time and the micro-climate of the bed, yielding insufficient variance in the values of t^m_μ to yield significant differences.

As an indicator of the impact of factoring continuous variables, the first two columns of table 7.5 use the age on a per year basis. In the final analysis age was refactored from a continuous variable to one with decadal bins to allow for the possibility that the significant differences as a function of age would be different at different decades. The impact of this decision on overall variance explained was very minimal. The adjusted partial R^2 reduced from 0.28 to 0.23. The overall R^2 reduced from 4.88 to 4.83. Refactoring variables is often be a trade–off between maximising variance explained and creating variables of the greatest use to the research design; in this case the decision did not make much difference.

Generally, sociodemographic variables explain less variance than physical ones (i.e. average external temperature and activity level), with the exception of BMI. The relatively high value of partial R^2 for BMI is notable and warrants further investigation (see 10.2). It is perhaps unsurprising that the explained variance of sociodemographic variables is low – the amount of time between the measurement of these variables and the physical variables was large, in some cases as many as 7 years may have elapsed between them. The fact that any variance is explained with these variables is therefore noteworthy, and may either point to the relative stability of these variables over time, or be a function of the large sample size of the study, or both.

It is important to caveat these findings with the observation that R^2 should only be interpreted alongside the other statistical measures used in this study. R^2 alone does not indicate a model's usefulness.

7.3.2 Significant variables

A related measure of the extent to which a model has explanatory power is the total number of significant variables that are predicted. This is shown for each metric in table 7.6

Metric	$p < 10^{-9}$	$p < 0.001$	$p < 0.01$	Total	Additional
t_{min}	15	6	4	25	4
t^m_{min}	12	10	1	23	3
t_{10}	17	2	5	24	4
t^m_{10}	13	3	7	23	0
t_μ	7	7	2	16	0
t^m_μ	6	6	1	13	1
t_{max}	5	12	5	22	1
t^m_{max}	6	5	6	17	1
t_{sd}	17	10	2	30	5
t^m_{sd}	20	8	3	31	4
t_{range}	18	7	3	28	5
t^m_{range}	16	9	2	27	3

Table 7.6: The number of statistically significant subcategory variables for each metric. The main portion of the table out of a possible total of 56 degrees of freedom used in chapter 7 (the total number of subcategories). The numbers of significant subcategories from the additional variables (*health satisfaction*, *financial satisfaction*, *heating type* is given in the final column, out of a additional total of 23. As mentioned in the text, these additional variable were only available for a subset of 37,730 participants.

Chapter 8

Results 4: Conditions associated with excess winter deaths

A stuffy room, with air constantly heated to 75°, is the most efficacious invention ever devised for ruining health.

Amelia E. Barr – Maids, Wives and Bachelors (1898)

8.1 Introduction

Excess winter deaths (EWDs) are defined as the total deaths occurring in winter over the non-winter average. It was shown in the literature review in chapter 2 how this definition does not necessarily adequately capture the relationship between ambient temperature and mortality. However, one useful result of calculating EWD is that it allows the conditions that ultimately lead to winter deaths to be identified. The Office for National Statistics produces yearly estimates of EWD, as well as their primary causes of death (ONS, 2018). It is these conditions, denoted C_{EWD}, and their prevalence, that will be the focus of this chapter.

8.2 Research question 2

"Are there associations between experienced temperature and the health conditions related to excess winter deaths (C_{EWD})?"

The simplicity of this research question allows it to be addressed using straightforward binomial regression. Beginning with t_{10}^{m} as the metric of interest of experienced temperature, the following regression equation was constructed (see section 4.7.2.3), and the effect estimates calculated. It tests associations between the variable C_{EWD} and the experienced temperature y of the i^{th} participant in the j^{th} regional centre.

$$\mathcal{L}(C_{EWD_{ij}}) = \beta_0 + y_{ij} + e_{ij} \tag{8.1}$$

As with the previous sections, the other lower metrics and the metrics of thermal variety can also be examined using this equation. Prior to this, multilevel structure is tested for to understand whether there is a strong effect of the region that participants live in on the prevalence of the C_{EWD}.

8.2.1 Multilevel structure

First, the value of LR is calculated in a similar manner as was done in chapter 6:

$$LR = 157.62 \quad (df = 1) \tag{8.2}$$

This is slightly stronger evidence of multilevel structure than for research question 3.4 and again the VPC is calculated:

$$VPC = 0.007 \tag{8.3}$$

Once again the degree of variance attributable to different regions is very small, so the simpler single level model was adopted. The geographical variation of the participants diagnosed with conditions relating to EWDs ($C_{EWD} = 1$) is shown in figure 8.1. The visualisation is consistent with the finding that participants for whom $C_{EWD} = 1$ does not exhibit clear variation as a function of centre location. For each centre, the number of participants for whom $C_{EWD} = 1$ and for whom $C_{EWD} = 0$, and the percentage incidence of C_{EWD}, is given in table 8.1. It is important to note that the variable *centre* only acts as a proxy for the geographical location of the participant, as figure 8.1 shows.

8.2.2 Single level binomial regression

The regression equation for this research question tests associations between the variable C_{EWD} in terms of log-risk and the experienced temperature t_{10}^{m} of the i^{th} participant, and has the following form:

$$\mathcal{L}(C_{EWD_i}) = \beta_0 + t_{10_i}^{m} + e_i \tag{8.4}$$

Practically, this was implemented in R using the following code:

```
glm(ewd~tm10, data=data,family=binomial(link=log),start=c(log(0.2)))
```

Specifying the start option equal to the log of the prevalence of C_{EWD} in the sample is required for model convergence. The parameter estimates are as follows:

$$RR(t_{10}^{m}) = 1.02[1.02 - 1.03] \qquad p = 1.92 \times 10^{-5} \tag{8.5}$$

$$RR(\beta_0) = 0.07[0.06 - 0.10] \qquad p < 2 \times 10^{-16} \tag{8.6}$$

The residuals of this model are shown in figure 8.2.

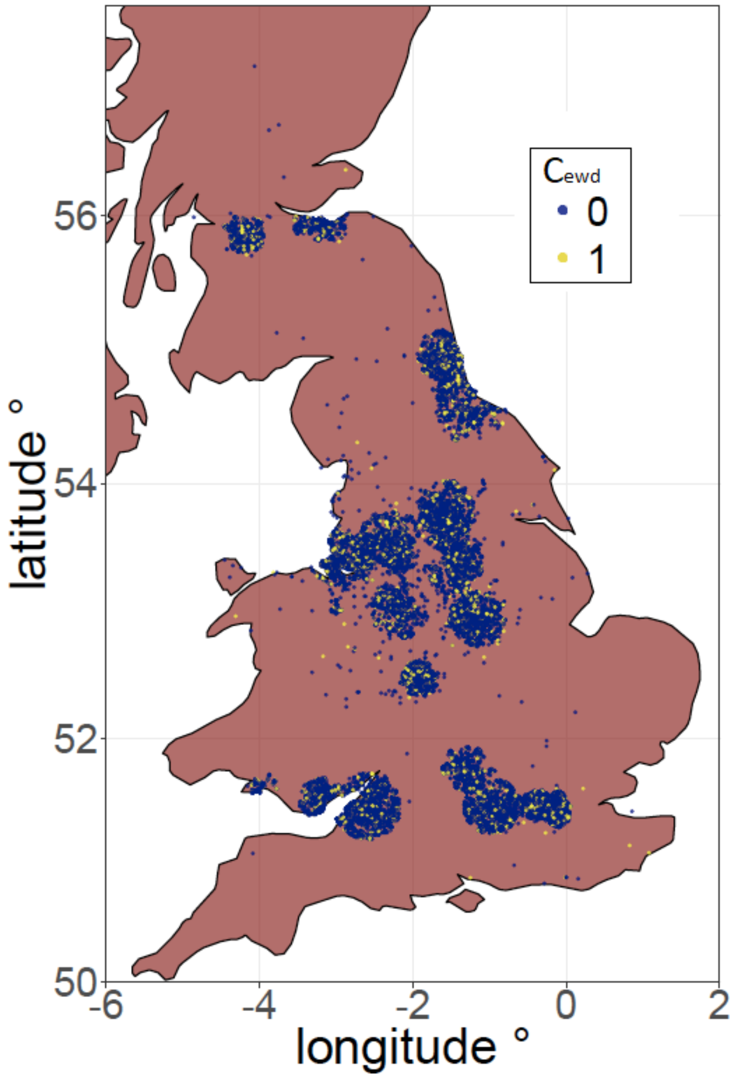


Figure 8.1: The geographical variation of the participants diagnosed with conditions relating to EWDs. $C_{EWD} = 1$ is given in yellow, $C_{EWD} = 0$ in blue. Table 8.1 gives the total number of participants in each centre who have $C_{EWD} = 1$

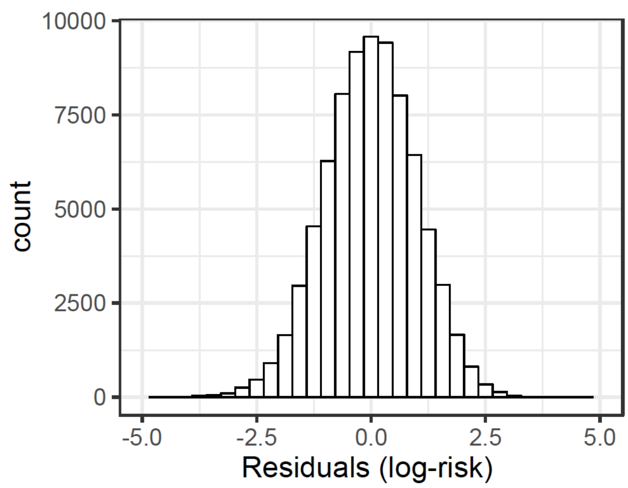


Figure 8.2: The residuals for the above fit of C_{EWD}. They are normally distributed. The standard deviation of the residuals is 0.87 in units of log-risk.

Centre	$C_{EWD} = 0$	$C_{EWD} = 1$	$\%(C_{EWD} = 1)$
Birmingham	3582	603	14
Barts	1894	245	11
Bristol	6835	914	12
Bury	2892	543	16
Cardiff	2097	349	14
Croydon	4533	631	12
Edinburgh	2581	347	12
Glasgow	1771	352	17
Hounslow	4727	639	12
Leeds	5475	883	14
Liverpool	3860	720	16
Manchester	1666	292	15
Middlesborough	2759	502	15
Newcastle	4091	867	17
Nottingham	4869	683	12
Oxford	2321	285	11
Reading	5040	688	12
Sheffield	4195	683	14
Stockport pilot	371	59	14
Stoke	1806	284	14
Swansea	314	51	14
Wrexham	86	18	17
Average	**3080**	**484**	**14**

Table 8.1: For each centre, the number of participants who do not have conditions associated with excess winter deaths, the number who do, and the percentage of the total who do

In binomial regression models the intercept does not have an interpretation. The value of 1.02 is interpreted as evidence that every °C increase in experienced temperature is associated with a 2% increase in the risk of C_{EWD}. However, as the previous chapter demonstrated, experienced temperature is known to have associations with other sociodemographic, housing and health variables. Therefore, the next section examines the risk of C_{EWD} with experienced temperature as well as the demographic and housing variables.

8.3 Research question 3

"Do combinations of sociodemographic factors, building factors and C_{EWD} have associations with low experienced temperature?"

As with research question 2, the appropriate regression to perform here is binomial. The regression equation combines the variables used in research questions 1 and 2. As before, the j subscript is dropped since the regional centres are not modelled:

$$\mathcal{L}(C_{EWD_i}) = \beta_0 + t^m_{10_i} + \sum_{k=2}^{9} \beta_k x_{k_i} + \beta_{13} x_{13} + \beta_{14} x_{14} + e_i \tag{8.7}$$

variables with a value of k of either 10, 11, or 12 correspond to the variables which are only available for a smaller subset of participants and are therefore not included at this stage.

8.3.0.1 Results

The full results of the regression model for research question 3 are given in table 8.2 and continued in table 8.3. Overall, the number of participants who had $C_{EWD} = 1$ was 10,523 (13.5%) and $C_{EWD} = 0$ was 67,239 (86.5%).

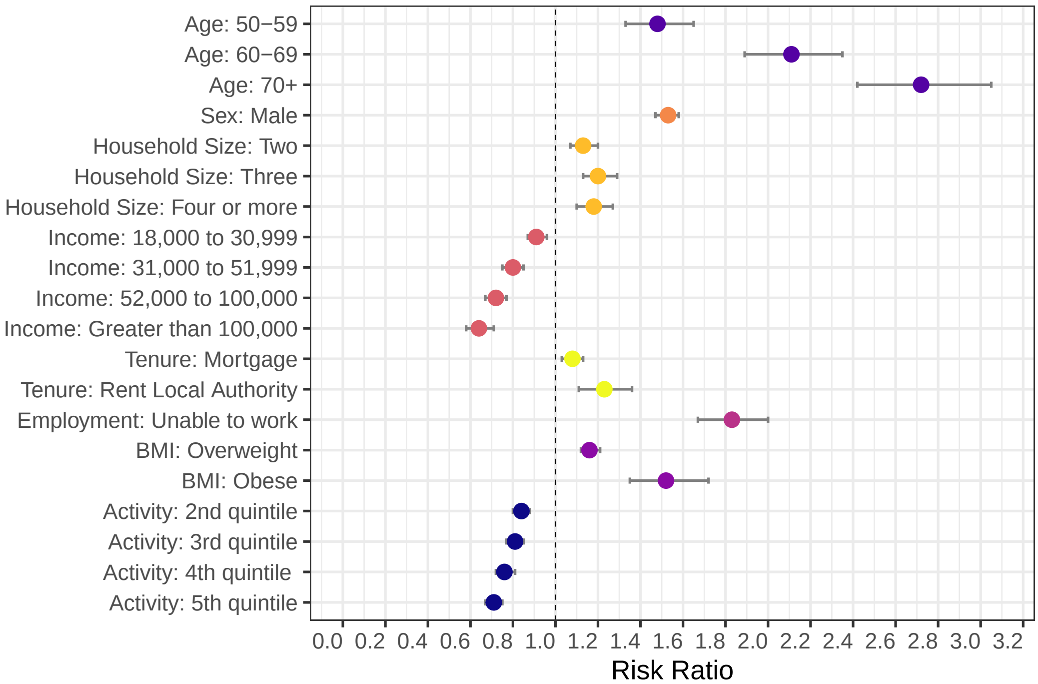


Figure 8.3: The explanatory variables which yielded significant risk ratio estimates for the sociodemographic and building factors in the model. A value of 1 is interpreted as having no impact on the risk of C_{EWD}. This figure does not include estimates for the experienced temperature or thermal variety. These are given below.

The most notable result of research question 3, relative to those of research question 2 described above, is that the association between C_{EWD} and the t_{10}^{m} experienced temperature is no longer significant after the inclusion of the covariates. The implications of this finding will be discussed further in chapter 9. However, a number of clear statistically significant results were evident from the regression, these are outlined as follows.

Age. Older participants were found to have a higher relative risk of C_{EWD}, those aged between 70–79 have a $RR = 2.72$ [2.42 – 3.05] (where square brackets denote the 95% confidence interval) over those aged 40–50. It is notable that the confidence intervals are particularly large for the age variable in comparison to the others. This is likely due to the variance of C_{EWD} within the variable itself.

Sex. Male participants had an $RR = 1.53$ [1.47 – 1.58] of having a condition associated

Continuous variables		**C_{EWD}: 0 Mean [sd]**	**C_{EWD}: 1 Mean [sd]**	**Risk ratio**
Experienced temperature t^{m}_{10} (°C)		28.6 [1.9]	28.7 [1.9]	1.00 [0.99 – 1.01]

Categorical variables	**subcategory**	**C_{EWD} : 0 n (%)**	**C_{EWD} : 1 n (%)**	**Risk ratio**
Age	40-49	5707 (93.9)	368 (6.1)	
	50-59	19364 (90.8)	1956 (9.2)	1.48 [1.33 – 1.65] ***
	60-69	30301 (85.6)	5106 (14.4)	2.11 [1.89 – 2.35] ***
	70+	11867 (79.3)	3093 (20.7)	2.72 [2.42 – 3.05] ***
Sex	Female	39155 (89.5)	4615 (10.5)	
	Male	28084 (82.6)	5908 (17.4)	1.53 [1.47 – 1.58] ***
Ethnic background	White	65135 (86.4)	10230 (13.6)	
	Mixed	357 (89.7)	41 (10.3)	0.99 [0.74 – 1.32]
	Asian or Asian British	556 (85.0)	98 (15.0)	1.16 [0.97 – 1.38]
	Black or Black British	524 (90.0)	58 (10.0)	0.90 [0.70 – 1.14]
	Chinese	144 (91.7)	13 (8.3)	0.86 [0.51 – 1.43]
	Other ethnic group	343 (86.8)	52 (13.2)	1.12 [0.88 – 1.44]
	Do not know	17 (85.0)	3 (15.0)	1.17 [0.41 – 3.38]
	Prefer not to answer	163 (85.3)	28 (14.7)	0.95 [0.68 – 1.33]
Household income	Less than 18,000	8663 (81.8)	1929 (18.2)	
	18,000 to 30,999	14972 (84.2)	2807 (15.8)	0.91 [0.87 – 0.96] **
	31,000 to 51,999	17491 (87.4)	2525 (12.6)	0.80 [0.75 – 0.85] ***
	52,000 to 100,000	15233 (89.5)	1788 (10.5)	0.72 [0.67 – 0.77] ***
	Greater than 100,000	4421 (91.2)	429 (8.8)	0.64 [0.58 – 0.71] ***
	Prefer not to say	4722 (86.2)	753 (13.8)	0.84 [0.77 – 0.91] **
	Do not know	1737 (85.6)	292 (14.4)	0.92 [0.82 – 1.03]
Tenure type	Own outright	37951 (85.2)	6586 (14.8)	
	None of above	244 (88.4)	32 (11.6)	0.92 [0.66 – 1.26]
	Prefer not to answer	184 (85.6)	31 (14.4)	1.08 [0.78 – 1.49]
	Mortgage	25313 (88.8)	3185 (11.2)	1.08 [1.03 – 1.13] *
	Rent Local Authority	1704 (81.3)	392 (18.7)	1.23 [1.11 – 1.36] **
	Rent private	1292 (86.3)	205 (13.7)	1.11 [0.98 – 1.26]
	Shared	145 (83.3)	29 (16.7)	1.47 [1.07 – 2.02]
	Rent free	406 (86.6)	63 (13.4)	1.04 [0.83 – 1.31]
Accommodation type	House/bungalow	61864 (86.5)	9690 (13.5)	
	None of above	73 (88.0)	10 (12.0)	0.80 [0.45 – 1.41]
	Prefer not to answer	10 (76.9)	3 (23.1)	1.19 [0.40 – 3.56]
	Flat	5245 (86.6)	813 (13.4)	0.98 [0.91 – 1.05]
	Temporary	47 (87.0)	7 (13.0)	0.81 [0.41 – 1.62]

Table 8.2: Results of the binomial regression of C_{EWD} with the demographic, housing and health factors described in the text. N=77,762. The total number of participants and the percentages for either $C_{EWD} = 0$ or $C_{EWD} = 1$ are given, along with the risk ratio and 95% confidence interval. The relative subcategory for each variable does not have an RR estimate. Significance levels: * $p < 0.01$, ** $p < 0.001$, *** $p < 1 \times 10^{-9}$.

Categorical variables	subcategory	C_{EWD} : 0 n (%)	C_{EWD} : 1 n (%)	Risk ratio
Employment status	In paid/self-employment	35511 (89.2)	4286 (10.8)	
	None of the above	299 (85.4)	51 (14.6)	1.10 [0.85 – 1.41]
	Prefer not to answer	89 (89.9)	10 (10.1)	0.76 [0.42 – 1.36]
	Retired	22700 (82.6)	4772 (17.4)	1.06 [1.01 – 1.11]
	Looking after home and/or family	2942 (90.9)	293 (9.1)	0.96 [0.86 – 1.08]
	Unable to work because of sickness or disability	1006 (71.3)	405 (28.7)	1.83 [1.67 – 2.00] ***
	Unemployed	794 (88.1)	107 (11.9)	0.85 [0.71 – 1.01]
	Doing unpaid or voluntary work	3239 (86.2)	520 (13.8)	1.06 [0.97 – 1.15]
	Full or part-time student	659 (89.3)	79 (10.7)	1.04 [0.84 – 1.27]
Fuel type	Gas hob or gas cooker	25237 (87.2)	3720 (12.8)	
	Gas fire	5405 (84.7)	974 (15.3)	1.05 [0.99 – 1.12]
	Open solid fuel (s.f) open fire	2039 (87.3)	296 (12.7)	1.01 [0.90 – 1.12]
	Gas hob & Gas Fire	17309 (85.7)	2879 (14.3)	1.05 [1.00 – 1.09]
	Gas hob & s.f. open fire	4003 (89.3)	478 (10.7)	0.92 [0.84 – 1.01]
	Gas fire & s.f. open fire	169 (86.7)	26 (13.3)	1.04 [0.73 – 1.48]
	Gas hob & Gas fire & s.f. open fire	837 (87.6)	119 (12.4)	0.98 [0.83 – 1.16]
	None of the above	12202 (85.8)	2019 (14.2)	1.01 [0.96 – 1.06]
	Prefer not to say	28 (75.7)	9 (24.3)	1.66 [0.96 – 2.88]
	Do not know	10 (76.9)	3 (23.1)	1.54 [0.64 – 3.73]
Body mass index	normal	27256 (89.2)	3306 (10.8)	
	underweight	430 (90.1)	47 (9.9)	1.00 [0.76 – 1.31]
	overweight	38758 (84.8)	6964 (15.2)	1.16 [1.12 – 1.21] ***
	obese	795 (79.4)	206 (20.6)	1.52 [1.35 – 1.72] ***
activity level during study week, by quintile, lowest to highest activity	1st quintile	12429 (80.4)	3034 (19.6)	
	2nd quintile	13346 (85.7)	2221 (14.3)	0.84 [0.80 – 0.88] ***
	3rd quintile	13581 (87.2)	1986 (12.8)	0.81 [0.77 – 0.85] ***
	4th quintile	13802 (88.6)	1776 (11.4)	0.76 [0.72 – 0.81] ***
	5th quintile	14081 (90.3)	1506 (9.7)	0.71 [0.67 – 0.75] ***
Household size	Single occupant	11147 (86.7)	1707 (13.3)	
	Two	2150 (27.2)	5755 (72.8)	1.13 [1.07 – 1.20] **
	Three	10568 (87.0)	1573 (13.0)	1.20 [1.13 – 1.29] **
	Four or more	13374 (90.0)	1488 (10.0)	1.18 [1.10 – 1.27] **

Table 8.3: Results of the binomial regression of C_{EWD} with the demographic, housing and health factors described in the text. N=77,762. The total number of participants and the percentages for either $C_{EWD} = 0$ or $C_{EWD} = 1$ are given, along with the risk ratio and 95% confidence interval. The relative subcategory for each variable does not have an RR estimate. Significance levels: * $p < 0.01$, ** $p < 0.001$, *** $p < 1 \times 10^{-9}$.

with excess winter deaths compared to female participants.

Household size. RR(C_{EWD}) also increased as a function of number of inhabitants in the home. Living in a home of 2, 3 or 4 and more, conferred an increased risk of C_{EWD}, of 1.13 [1.07 – 1.20], 1.20 [1.13 – 1.29] and 1.18[1.10 – 1.27] respectively. This relationship was not monotonic, since the confidence intervals of the categories 3 and 4 or more overlap.

Household income. In contrast to the findings reported in the previous chapter, there was a clear income effect for C_{EWD} risk. Risk reduced monotonically as a function of income. Those living in homes which earned more than £100,00 had $RR = 0.64$ [0.58 - 0.71] relative to those earning less than £18,000 a year.

Tenure type. Relative to owning a home outright, participants with a mortgage had increased risk of C_{EWD}, $RR = 1.08$ [1.03 – 1.13], as did those who rent from the local authority $RR = 1.23$ [1.11 – 1.33].

Employment type. The only significant difference in risk of C_{EWD}, relative to those in paid work or self-employed, was that those who were unable to work because of sickness or disability had a much increased risk of C_{EWD} $RR = 1.83$ [1.67 – 2.00].

BMI. Body mass index was found to have a significant association of C_{EWD}. Obese participants had $RR = 1.52$ [1.35 – 1.72] above participants of normal weight.

Activity level. Participants who recorded the highest quintile of activity had in decreased risk of C_{EWD}, $RR = 0.71$[0.67 - 0.75]

The inclusion of the variables *heating type*, *financial situation satisfaction* and *health satisfaction* resulted in problems with the model converging. This is likely due to a combination of the increased number of variables and the decreased number of participants that were available for these variables. A discussion of these variables is reserved for subsection 8.4.2.

8.3.1 Statistical checks

Like the multiple linear regression model used in the previous chapter, binomial regression also has a number of statistical requirements in order to constitute a valid regression. The first check is on the residuals of the model - these are shown in 8.4. They are sufficiently normal to constitute a valid model.

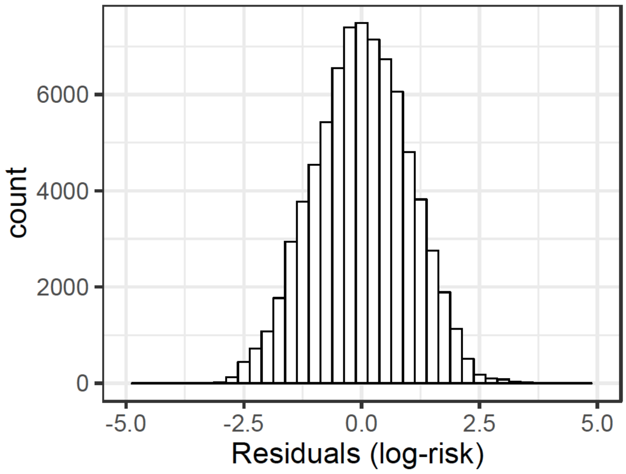


Figure 8.4: The residuals for the above fit of C_{EWD} which includes t_{10}^{m}. They are normally distributed, with a slight negative skew which is insufficient to cause problems. The standard deviation of the residuals is 0.85 log-risk.

Next, the issue of multicollinearity was tested. All values of the VIF were below 2, so it can be concluded that there is minimal chance of multicollinearity affecting the model.There is a similar requirement on linearity as in the previous models. External temperature is not included in this regression since there is no conceivable mechanism by which the external temperature could be associated with C_{EWD} because the participants received the AX3 at a random time. Therefore, the only continuous variable that needs to be considered is t_{10}^{m}. This is shown in figure 8.5. The programming language R which was used to construct the plot was unable to process the entire sample of 77,762 participants so a random sub-sample of 20,000 participants was used to construct the test. This was carried out several times to ensure that the nature of the plot was not a result of the particular random sample selected.

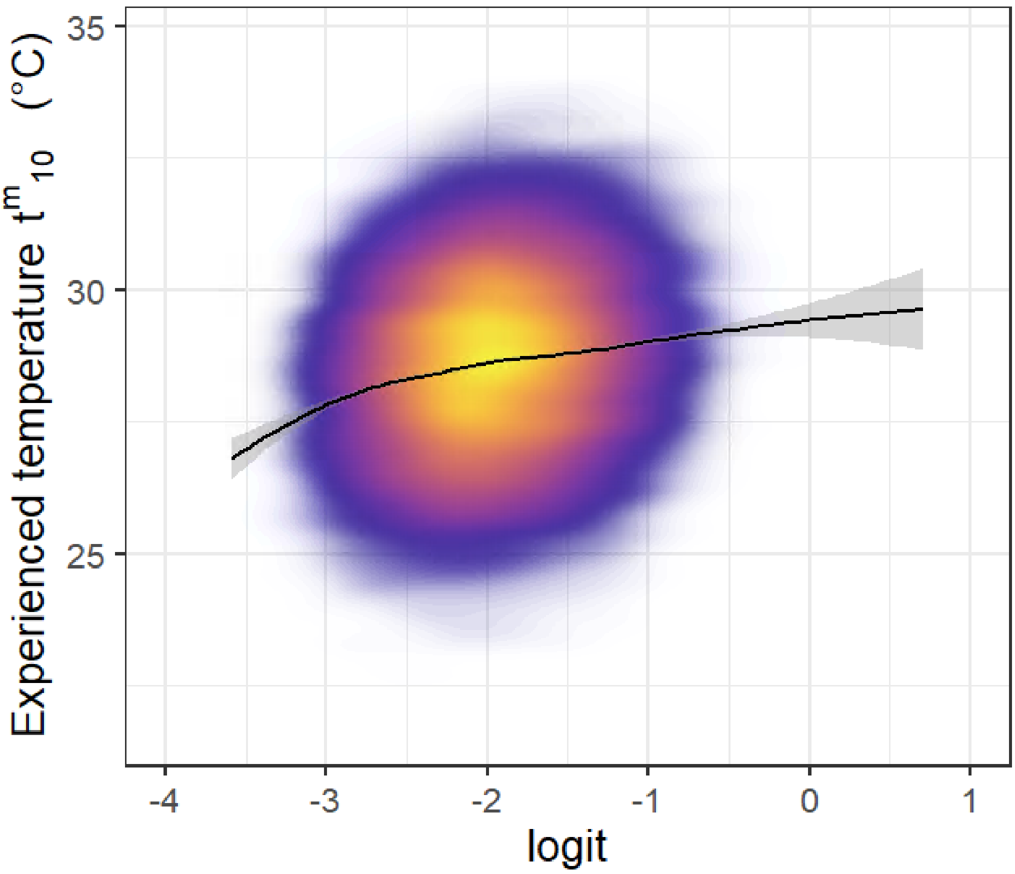


Figure 8.5: The relationship between the predicted logit value (see section 4.7.2.3) of the model and the value of t_{10}^{m}. The relationship is linear, with slight deviations from linearity at the extreme values, as shown by the LOESS fit. The large amount of spread accounts for the weakness of the relationship between t_{10}^{m} and C_{EWD}. As in previous figures, the density of points is shown to represent the 20,000 participants used to create the figure.

Finally, for binomial regression, it is advisable to ensure that no one value or set of values overtly impacts the regression estimates (Kassambara, 2018). For this, the model is plotted in terms of the Cook's distance (Aguinis et al., 2013), as is shown in 8.6. A number of outliers are visible in this plot, however, to calculate the impact of these outliers the standardised residuals are calculated. For this, should any have a value greater than 3 then they must be investigated further (this rule of thumb derives from the observation that 99.9% of all standardised residuals should be within ±3.29 of the mean).

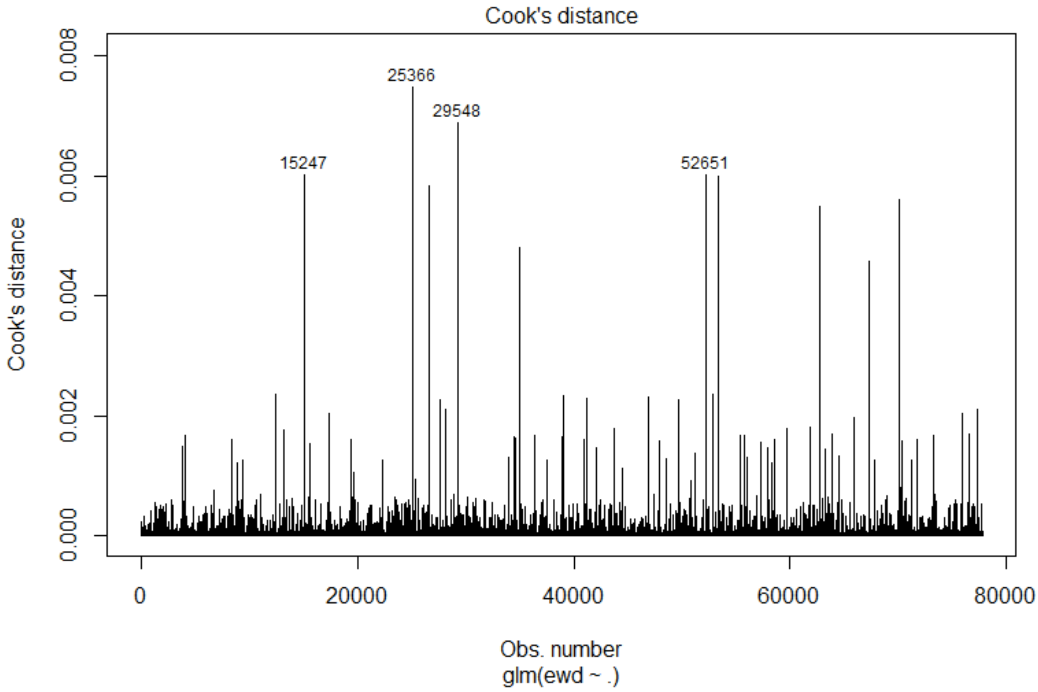


Figure 8.6: Cook's distance for each participant in the model.

Computation of the standardised residuals reveals none with a value greater than 3. It is therefore possible to conclude that the model is not adversely impacted by outlying values.

8.3.2 Robustness

As chapter 6 showed, t_{10}^m, which is the first decile of experienced temperature for the times at which the participant was likely sedentary, varies with sociodemographic and housing factors, but most variance is explained by the external temperature variable. Since there is no means by which the external temperature at the time of the AX3 measurement could impact C_{EWD} it was not included in the model. However, t_{10}^m is known to vary with external temperature, which could impact the results indirectly. Therefore, a standardised t_{10}^m was computed, for which the external temperature was removed. This was done by using the predicted value of t_{10}^m from the regression given in figure 6.3 and subtracting this value from the t_{10}^m values. This value of t_{10}^m was then used in the binomial regression above, but there was no difference in the regression estimates. This is likely explained by the fact that the overall variance of t_{10}^m explained by external temperature is less than 4%, $R^2 = 0.039$.

8.4 Other metrics of experienced temperature and thermal variety

Metric	RR without	p value	RR with covariates	p-value
t_{10}	1.04 [1.03 – 1.05]	≈ 0	1.01 [1.00 – 1.02]	0.04
t_{10}^m	1.02 [1.02 – 1.03]	≈ 0	1.01 [0.99 – 1.01]	0.0448
t_{min}	1.04 [1.03 – 1.04]	≈ 0	1.01 [1.01 – 1.02]	0.0001
t_{min}^m	1.04 [1.03 – 1.04]	≈ 0	1.01 [1.00 – 1.01]	0.014
t_{μ}	1.02 [1.01 – 1.03]	≈ 0	1.00 [0.99 – 1.01]	0.632
t_{μ}^m	0.99 [0.98 – 1.00]	0.021	1.00 [0.99 – 1.01]	0.72
t_{sd}	0.82 [0.79 – 0.84]	≈ 0	0.95 [0.94 – 0.98]	0.001
t_{sd}^m	0.82 [0.79 – 0.84]	≈ 0	0.97 [0.94 – 1.00]	0.04
t_{range}	0.96 [0.96 – 0.97]	≈ 0	0.99 [0.99 – 1.00]	0.001
t_{range}^m	0.96 [0.95 – 0.96]	≈ 0	0.99 [0.99 – 1.00]	0.02
t_{iqr}	0.92 [0.91 – 0.93]	≈ 0	0.98 [0.97 – 1.00]	0.02
t_{iqr}^m	0.92 [0.91 – 0.94]	≈ 0	0.98 [0.98 – 1.01]	0.3

Table 8.4: n=77,762 for both model types. A value of $p \approx 0$ denotes a value less that 10^{-16}

The above results only considered the use of t_{10}^m and its association with C_{EWD}. The present section considers the other measures of experienced temperature and thermal variety used in previous chapters. The estimates of the risk ratio (RR) associated with each variable are given in 8.4. This includes both the RR for the model using only the experienced temperature (or thermal variety) metric as well as the model which includes covariates. This is equivalent to differences between the models for research question 2 and research question 3. With reference to research question 2, it is instructive to include the risk of C_{EWD} without any covariates. It is helpful to know the uncorrected risk associated with a particular factor, and to compare this with the estimates including likely confounders. It shows that without confounders the risk of C_{EWD} reduces with increased thermal variety. This is the case under for the metric derived from the range as well as the standard deviation. This result also holds for the interquartile range, which is discussed more below in section 7.1.4 on alternative metrics. The addition of the confounding variables shifts all estimates for the risk associated with the various metrics used in this chapter towards one. t_{sd} and t_{range} are the only metrics which remain significant. It may be possible to conclude from this that whatever risk reduction that is associated with thermal variety is not conferred at times at which the participant is sedentary. It does not follow that the participant is necessarily outdoors during these times.

The risk ratio (RR) estimates for the sociodemographic variables for the other models which use different metrics are so similar that they are not reported explicitly. However, since the RR estimates for the models using t_{sd} are the greatest in magnitude they are reproduce in the appendix (tables B.9 and B.10). The thermal variety metrics t_{sd} and t_{sd}^m are the focus of the results presented in this section for the same reason. The estimates for

t_{10}^m result in RR greater than one. For t_{sd} and t_{sd}^m are less than one. This is interpreted as suggesting that for each degree the thermal variety increases, the risk of C_{EWD} reduces. Since this is the opposite direction of the model t_{10}^m the statistical check will also be reported for the variety metrics. However, in general, the robustness checks which held for t_{10}^m also hold for the other metrics.

Focusing on the model using t_{sd}, the primary result is that increased thermal variety is associated with a decreased risk of C_{EWD}. For a model using t_{sd} alone, the estimates are $RR = 0.82\,[0.79 - 0.84]$. This is interpreted as a 18% reduction in risk of C_{EWD} for every degree that the t_{sd} thermal variety increases. Following the inclusion of covariates, this estimate reduces to $RR = 0.95\,[0.94 - 0.98]$. Of course, as with all the associations highlighted in these results chapters, the causal nature of this relationship cannot be inferred from this analysis alone. This issue will be further addressed in the discussion in chapter 9.

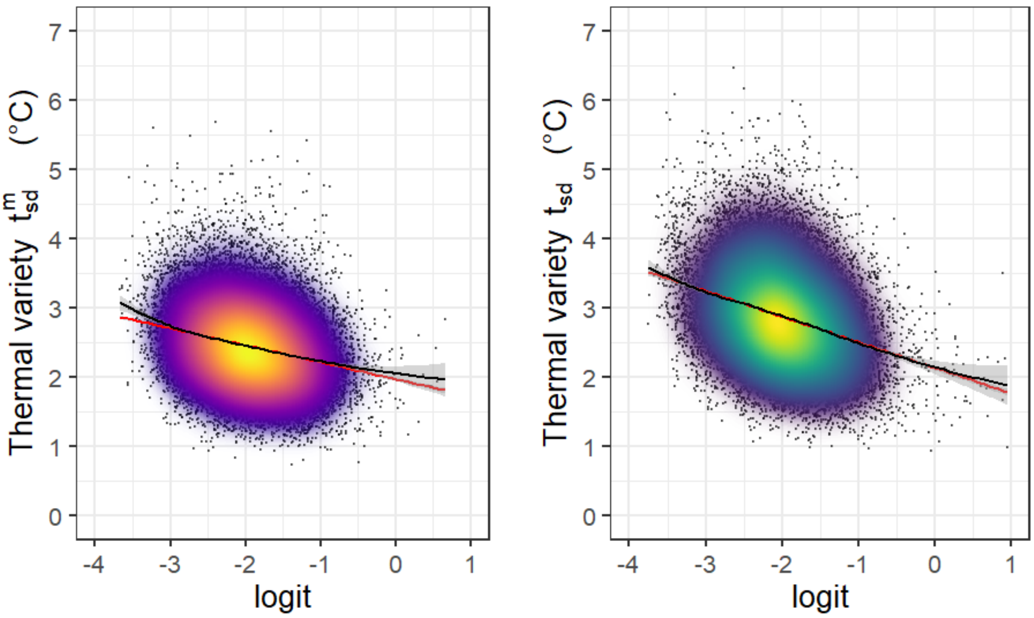
As mentioned above, this model also found that the risk of C_{EWD} reveals very similar numerical estimates for the potentially confounding variables, which suggests that the impact on the model of including t_{sd} instead of t_{10} are minimal; the risk ratio estimates of each demographic, building or health factor differ between the two models by no more than 0.02. This is likely because the experienced temperature contributes little overall to the risk of C_{EWD}, even though a significant relationship is observed in the case of t_{sd}. Likewise, the use of the other metrics also did not modify the estimates for the RR for the sociodemographic and building covariates.

8.4.1 Statistical checks

As with the statistical check reported above for the t_{sd}^m mode, the plot of the logit as a function of both t_{sd} and t_{sd}^m is given in figures 8.7a and 8.7b. Tests to see if the model was adversely impacted by one particular value revealed that this was not the case. Again, the VIF was calculated and no value was above 3, suggesting multicollinearity is not an issue. It is therefore possible to conclude that binomial regression used for these variables is valid. Similar tests on the other variables were also conducted and showed the models using the other metrics were not impacted negatively by either non-linearity or extreme values.

8.4.2 Additional variables

It was alluded to above that including the extra variables *health satisfaction*, *heating type* and *financial situation satisfaction* prevented the model from converging. This was likely because the number of variables was larger, and the number of participants smaller than the previous model. This problem was addressed by removing the variables *ethnicity*, fuel type and *accommodation*, for which there were no significant relationship with risk of C_{EWD}. Following this, each of the additional variables were added to the new model in turn. The only model which converged and showed significant results was one including



(a) The relationship between the predicted logit value of the model and the value of t_{sd}^{m}.

(b) The relationship between the predicted logit value of the model and the value of t_{sd}.

Figure 8.7: Both relationships are linear, with slight deviations from linearity at the extreme values in figure 8.7a for t_{sd}^{m}, as shown by the LOESS fits (see section 4.7.2.3). The relationship for t_{sd} is steeper than for t_{sd}^{m}. This accounts for the greater magnitude risk ratio estimate for t_{sd}. LOESS regression is in black. Simple least squares, which is used by the model, in red. As in previous figures, the density of points is shown at regions where there is high density of point.

health satisfaction. These results are given in table 8.5. Compared to the full model, the most notable change is that the estimate for t_{sd} became non-significant. This is likely due in part to the introduction of the *health satisfaction* variable – although it should be noted that the relationship between these variables is minimal; a regression of *health satisfaction* alone on t_{sd} accounts for only 0.08% of the variance of t_{sd}. This means that the issue of co-linearity is unlikely to be the reason for the change. A second factor is the reduced sample size of 38,003. The most likely explanation is that the specific way that *health satisfaction* interacts with other regression variables is the reason. The other changes are now examined in detail.

First, tenure type also becomes non-significant. Again, this is not to say that the relationship revealed by the previous model is erroneous, merely that inclusion of *health satisfaction* modifies the results. The risk of C_{EWD} associated with age become greater in magnitude in the model with extra variables. For example, those aged over 70 years have a C_{EWD} risk of 3.36 [2.85 − 3.95] (up from 2.72 [2.42 − 3.05]). However, the fact that the confidence intervals of the new estimates overlap with the previous model's estimates suggests that the addition of health satisfaction only has a moderate impact on the model.

The risk ratio associated with the variable sex does not change, outside the limits of the confidence intervals. The risk ratio of C_{EWD} associated with those who are unable to work goes down from $1.82\,[1.66-1.99]$ to $1.28\,[1.11-1.46]$. The risk associated with living in a household which earns more than £100,000 per year changes from $0.64\,[0.57-0.71]$ to $0.70\,[0.61-0.82]$. As part of this, the risk associated with earning between £18,000 and £30,999 becomes non-significant. Likewise, there is a loss of significance for obese participants. The risk associated with being overweight remains at $1.11\,[1.05-1.17]$ (cf. $1.15\,[1.11-1.20]$ in the previous model). For the activity variable, the risk associated with being in the highest quintile changes from $0.72\,[0.68-0.77]$ to $0.82\,[0.75-0.90]$. Similarly, other activity quintiles also show reduced magnitude of risk in the new model. The estimate of the risks as function of number in the household does not change. Finally, there is a very strong effect with health satisfaction; those who are extremely unhappy have a risk of $3.30\,[2.63-4.14]$ compared to those who are extremely happy. Overall, these finding underlies the importance of building regression models which have specific research questions in mind. It is not surprising that those who have diagnosed illnesses would be less satisfied with their health – in this sense there is a degree of circularity if health satisfaction is included. For this reason the remainder of the discussion will not include the health satisfaction variable in the model of C_{EWD} risk. However, it is useful to note that *health satisfaction*, which is a subjective measure of health, is strongly associated with a measure (C_{EWD}) determined by physician diagnosis, which is less subjective.

8.5 Summary

This chapter has addressed the second and third of the three main research questions given in chapter 4. For the second research question "are there associations between experienced temperature and the health conditions related to excess winter deaths (C_{EWD})?", it was hypothesised that "those who have health conditions associated with excess winter deaths will be more likely to have higher experienced temperature". The results presented in table 8.4 show that all metrics had a significant relationship with C_{EWD} and that higher temperatures were associated with an increased risk of C_{EWD}. This finding is consistent with the evidence that was reviewed in the literature that those with long term disabilities were found to live in warmer homes than those who did not have long term disabilities (Huebner et al., 2018). These results also held for metrics which measured thermal variety. For these metrics, higher thermal variety was associated with a decreased risk of C_{EWD}. Although the casual relationships cannot be addressed in a simple regression such as this, this finding is consistent with picture that those who have conditions associated with excess winter deaths have less diverse environments than those who do not have conditions associated with excess winter deaths. These findings are considered further in the following discussion chapter.

Explanatory variable (relative subcategory)	Subcategory	Risk ratio (C_{ewd})
Thermal variety t_{sd} (°C)	-	0.97 [0.94 – 1.01]
Age (40 - 49)	50-59	1.61 [1.38 – 1.88]***
	60-69	2.59 [2.22 – 3.01]***
	70-79	3.36 [2.85 – 3.95]***
Sex (Female)	Male	1.50 [1.42 – 1.58]***
Income (less than 18,000)	18,000 to 30,999	0.96 [0.89 – 1.03]
	31,000 to 51,999	0.86 [0.79 – 0.94]**
	52,000 to 100,000	0.73 [0.66 – 0.80]***
	Greater than 100,000	0.70 [0.61 – 0.82]**
	Prefer not to say	0.83 [0.74 – 0.93]*
	Do not know	0.89 [0.76 – 1.04]
Tenure (Own outright)	None of above	0.82 [0.52 – 1.30]
	Prefer not to say	1.42 [1.00 – 2.02]
	Mortgage	1.08 [1.02 – 1.15]
	Rent Local Authority	1.13 [0.99 – 1.30]
	Rent private	1.05 [0.88 – 1.24]
	Shared	1.41 [0.95 – 2.09]
	Rent free	1.08 [0.78 – 1.51]
Household size (single occupant)	Two	1.19 [1.11 – 1.28]**
	Three	1.23 [1.12 – 1.35]**
	Four or more	1.24 [1.12 – 1.37]**
Employment status (In paid employment or self-employed)	None of the above	1.04 [0.72 – 1.50]
	Prefer not to answer	0.43 [0.14 – 1.31]
	Retired	0.99 [0.92 – 1.06]
	Looking after home and/or family	0.89 [0.75 – 1.05]
	Unable to work because of sickness or disability	1.28 [1.11 – 1.46]**
	Unemployed	0.60 [0.46 – 0.80]**
	Doing unpaid or voluntary work	0.92 [0.81 – 1.04]
	Full or part-time student	1.13 [0.84 – 1.52]
Body Mass Index (normal)	Underweight	1.02 [0.70 – 1.47]
	Overweight	1.11 [1.05 – 1.17]**
	Obese	1.09 [0.91 – 1.31]
Health Satisfaction (Extremely Happy)	Very happy	1.11 [0.96 – 1.29]
	Moderately happy	1.64 [1.42 – 1.89]***
	Moderately unhappy	2.48 [2.13 – 2.89]***
	Very unhappy	3.52 [2.96 – 4.19]***
	Extremely unhappy	3.30 [2.63 – 4.14]***
	Prefer not to answer	2.05 [0.58 – 7.26]
	Do not know	1.85 [1.13 – 3.02]
Activity quintile (1^{st})	2^{nd}	0.90 [0.84 – 0.97]*
	3^{rd}	0.90 [0.83 – 0.96]*
	4^{th}	0.87 [0.80 – 0.94]**
	5^{th}	0.82 [0.75 – 0.90]**

Table 8.5: The estimates for a reduced model which successfully converges. It includes the variable of *health satisfaction*. It is not taken further since the inclusion of *health satisfaction* is somewhat circular. Significance levels: * $p < 0.01$, ** $p < 0.001$, *** $p < 1 \times 10^{-9}$. N=38,003.

For research question three, “Do combinations of sociodemographic factors, building factors and C_{EWD} have associations with low experienced temperature?” the hypothesis was that “those who have health conditions associated with excess winter deaths will not be more likely to have higher experienced temperature if they are also in low income households.” The approach taken to answer this question was to include covariates in the binomial regression model for C_{EWD}. The research question has been answered in the affirmative by the models constructed here – the metrics t_{10}, t_{min} and t^{m}_{min} retained significant relationships with C_{EWD} once covariates were introduced into the model. These covariates, such as age, sex and income, contributed far more to overall model than the metrics of experienced temperature. The metrics of thermal variety t_{sd}, t^{m}_{sd} and t_{range} also retained significant relationships once covariates were introduced. The clearest effect was for the t_{sd} metric.

Focusing specifically on the income effect, relative to the poorest households, the risk of C_{EWD} steadily reduces with increasing income. This is not the same as suggesting that higher income guards against the effects of cold related C_{EWD} – such an observation is outside the scope of the research design – but it is consistent with the conceptual model that higher income households may have more means at their disposal to avoid C_{EWD}. Again, the findings are discussed further in following chapter.

Chapter 9

Discussion

Our apartment with central heating has turned out to be too warm - although it may be because of the warm weather we are having here

V. I. Lenin - A letter to his mother (November 4, 1909)

This chapter critically discusses the results presented in chapters 6, 7 and 8. In light of these results it then reviews the thesis as a whole and considers the potential implications of the findings.

9.1 Results

The first research question of this thesis, set out in chapter 3, was "Does experienced temperature vary with sociodemographic and building variables [e.g. sex, age, ethnicity, income, building type, tenure]?". The general answer to this question is that variation is observed, but the specifics of the observations provide the most interesting aspects of the results.

Returning to the hypotheses and dividing them up will help structure the following discussion. It was hypothesised that no significant differences in experienced temperature would be measured as a function of sex, ethnicity and income. This was done on the basis of the absence of evidence, or contradictory evidence, in the literature regarding the variation of wrist temperatures and dwelling temperatures as a function of these variables. In terms of the lower metrics, a moderate and consistent effect was observed which showed that males experience colder temperatures than females. This may be explicable from a thermal comfort prospective which suggests that men prefer cooler environments for metabolic reasons, as was explored in chapter 6.

The results for ethnicity were somewhat inconclusive. An absence of theoretical or literature based reasons for these observed differences, such as those of a Chinese ethnic background have a significantly lower t_{10}^{m} than White participants, prompted speculation regarding culturally specific heating practices, but drawing relationships between ethnic background and cultural practice is almost always problematic and usually incorrect.

The results for income were inconsistent. The finding that t_{min} and t^{m}_{min} decrease with increasing income was not observed in the first decile metrics. There was some evidence from the measures of thermal variety, specifically the difference between the standard deviation, range and interquartile range, which suggested that higher income participants might have an experienced temperature distribution which is narrower around the mean and wider at the extremes. This could indicate that the capacity to control the circumstances of thermal experience increases with income. This would require more investigation.

Hypothesis 1, that experienced temperature would increase as a function of increasing age and decreasing health satisfaction, was born out by the findings. This was particular backed up by the results using the thermal variety metric. The final portion of the hypotheses, that those living in Local Authority housing would have warmer lower metric experienced temperatures was shown in all metrics. This accords with the literature which finds that Local Authority housing is constructed to a higher standard than privately owned stock (Hamilton et al., 2017). This result may also go some way to explaining the inconsistent findings as a function of income – there is degree of heterogeneity in the housing quality of lower income households in the UK as not all low income families live in social housing. At the same time, a proportion medium and high income families live in housing that was once socially owned (Palmer and Cooper, 2013).

In order to examine the other findings more closely, the following section splits the discussion into two parts. The first addresses the lower metrics (t_{min} and t_{10}) and the second considers thermal variety (t_{sd}). The previous three chapters have shown that the picture of the data produced by the regression models depends to an extent on the choice of metric. The lower metrics tend to produce consistent results, and the thermal variety produce a slightly different set of results. In this sense, each acts as a lens which emphasises or occludes different parts of this picture. The question of what extra insight is gained by restricting the analysis to sedentary time periods is addressed in the final part of this section. Forming an overall picture of the relationship between experienced temperature and the sociodemographic, building and health is the primary goal of the first portion of this chapter.

9.1.1 Lower metrics

Both t_{min} and t_{10} give a similar picture of the underlying data for the majority of the variables considered in this study. In the Pre–analysis plan (PAP) it was expected that the t_{min} would be too prone to the effect of outliers to produce coherent results. While this likely the case at the individual level, averaging the effect across multiple participants reduces this risk. One of the strengths of multiple linear regression is the ability to create aggregate estimates of effects. However, it must be born in mind that the minimum experienced temperature still only corresponds to a single minute's worth of exposure.

The primary variable which explains the observed variance in both these metrics is the average external temperature. From a perspective of basic thermodynamics this is perhaps unsurprising, but an alternative emphasis is useful. In a fundamental sense it suggests that the temperatures people experience vary with season. This accords with the picture we have of the that the built environment, but it further suggests that clothing choices, social practices, and the host of other processes that construct everyday life in the UK do not prevent the wrist-worn AX3 sensor from detecting the seasonal variation in temperature. At the outset of this study it was not at all clear that the sensor would be sufficiently sensitive to do this. A further point is worth emphasising. The finding that winter experienced temperatures are colder than summer also points to a real basis for the phenomenological experience of winter – it is not simply the case that winter feels colder, it is also measurably so at the level of the individual. This latter point is crucial. Despite the apparent banality of this finding, no other study to date has measured the winter drop in temperature in a population at the level of individual experienced temperatures.

Of course, the most important aspect of the findings here is the differential effects of this cold exposure across different types of people. In this regard, the lower temperature metrics present a generally coherent picture. Exceptions to this coherence will be noted below. The clearest effect is related to the activity level of the participant. More active people have lower experienced temperatures. Since the multiple regression model balances the effect estimates across all the explanatory variables, it is possible to distinguish effects that are happening at the same time as this activity effect. Separately, therefore, it distinguishes that as people age, they tend to experience less cold temperatures. At the same time, an effect as a function of participant health is also clear – those who are extremely unhappy with their health tend to have warmer experienced temperatures than those who are extremely happy. This effect is less strong than either the activity or ageing effects. A clear effect is seen for those who are unable to work due to sickness of disability compared to those in employment. They tend to have warmer experienced temperatures, according to the lower metrics. This may point to desire for warmer temperatures to alleviate the impacts of the health conditions, or the limitations on mobility caused by the health conditions themselves reducing the chance of cold exposure.

The study shows that males have moderately colder experienced temperatures, as measured by the lower temperature metrics. These effects are smaller than those associated with difference between winter and summer, or being in the youngest group versus the oldest, for example. The relative difference for sex is around the same as the difference for a decade of ageing. However, males also have a lower mean experienced temperature, especially during sedentary periods.

A clear effect is observable in comparing households that have a solid-fuel open fire for heating in winter, versus those who do not. As the chapter 8 laid out, this may be due to the typical location of homes with open fires, as well as the building fabric and the heat distribution within the home, but assessing this is not possible with the available variables. There are no significant differences in this variable for the mean experienced temperature.

A consistent effect is observed for those who are overweight relative to those whose BMI is normal, they have warmer experienced temperatures according to the lower temperature metrics. The magnitude of this effect is slightly smaller than the one observed for sex difference. However, the opposite effect is seen for the mean experienced temperature. Inconsistent results are found for those who are obese when only sedentary time periods are considered, but using all time periods shows that obese participants are even warmer than overweight participants are found to be. Again, these findings should be interpreted alongside the mean experienced temperature, which are significantly lower than normal BMI participants. Taken together, these results point to a narrowing of the thermal variety experienced as a function of BMI, which will be discussed below. A study due to Mavrogianni et al. (2013) suggested that the link between reduced cold exposure and increased obesity incidence is plausible. However, the uncertainty in measures of domestic temperature makes determining whether this relationship has changed over time difficult.

Focusing on the activity, age and health factors allows the following summary. Colder experienced temperatures are associated with those who are more active, feel more healthy and are younger. This result also holds, with exception of health satisfaction, for times at which the participant is sedentary. For health satisfaction during sedentary times, the results are not statistically significant, but the effect size estimates do not contradict this picture. In the next section, the clearest findings from the portion of the study which used thermal variety are reviewed.

9.1.2 Thermal variety

The previous chapter defined the thermal variety as the standard deviation of experienced temperature. It showed that the variance explained by the explanatory variables was highest with t_{sd} as the outcome variable. It is preferable compared to the experienced temperature range since the upper experienced temperatures likely correspond to short term exposure to warm water during bathing. The interquartile range only characterises the central portion of the variation.

The findings using the thermal variety metric complement the above findings regarding lower temperature exposure. There is a clear reduction in thermal variety as a function of age; older participants have lower thermal variety than younger participants.

There was no effect observed as a function of income, but this was likely because an increasing thermal range as income increased was offset by a reduced interquartile range. This may suggest that richer participants spend the majority of their time in a reduced range of thermal environments, but take occasional excisions into more thermally diverse environments. This could also be linked to increased showering or bathing rates amongst the participants who live in higher income households.

Small effects were observed for male participants, whose thermal variety was found to be slightly less than females. They were found to have an increased range and a reduced interquartile range, in a similar manner as the income effect, although this relationship did not hold for when the participants were likely sedentary.

In agreement with the findings on the lower temperature metrics, the presence of an open solid-fuel fire led to increased thermal variety. Again, this may be due to the home location, as well as building fabric and home heat distribution factors.

Higher thermal variety was observed for the more active participants, even for times when they were likely sedentary. This accorded with the finding for health satisfaction of the participant. Thermal variety reduced as participants became less satisfied with their health.

9.1.3 Are both metrics necessary?

The lower metrics of experienced temperature (t_{10}, t_{10}^m, t_{min} and t_{min}^m) demonstrated clear relationships with several of the sociodemographic, building and health factors used in the stud. Despite differences between the metrics, they generally showed that participants who were younger, healthier, more active and who had greater health satisfaction tended to have increased exposure to cold. The results for the metrics of thermal variety were broadly consistent with this. A natural question that therefore arises at this stage is whether both kinds of metric, i.e. both lower temperature and thermal variety, are required for an understanding the relationships between the temperatures people experience and their sociodemographic status and health factors. They produce complementary but slightly different pictures of the data. The results using thermal variety tend to be more consistent, and a greater number of the explanatory variables are found to be statistically significant. The reason for this is two–fold. The standard deviation takes into account every value in the time-series when it computed. This contrasts the minimum and first decile metrics which only incorporate a portion of the time-series into their calculation. In this sense there is *more* data which goes into the standard deviation. Perhaps, more importantly, the standard deviation is, to an extent, independent of systematic sensor error. Two sensors can differ in their recording of the temperature by as much as 2°C. This offset would be present in the mean, minimum, or first decile measurements. For the standard deviation this would not be the case, since it characterised by the root mean squared deviation from the mean. Assuming the width of a degree is measured consistently across devices (which given the precision of 0.3°C is likely) then the standard deviation would be less prone to error than the other metrics. A similar argument holds for the range of experienced temperatures. The values of the unaffected metrics would them be less washed out, compared to those metrics which are impacted by sensor error. Of course, at the limit of a high number of readings, the errors are assumed to be uniformly distributed, and so would cancel out even for the minimum and lower temperature metrics, and so this explanation probably only contributes in part to the observed difference. However, it is also the case that the standard deviation *uses* readings from the higher experienced temperatures of the participant. Therefore, the thermal variety may be influenced by the impact of increased showering to a greater extent than the lower metrics are. The lower metrics are more likely to capture cold exposure at the expense of explained variance. The thermal variety on the other hand has more explanatory power, but it less certain that this explanation has to do with cold exposure per se, as it could be the result of both cold and warm exposure.

9.2 Conditions associated with excess winter deaths

Research question 2 and 3 sought to understand the impact of experienced temperature on the risk of having a disease associated with excess winter deaths (C_{EWD}). It was hoped that estimating the risk factors associated with C_{EWD} would shed light on who might be most vulnerable to the impacts of cold. Even though cold is not thought to cause C_{EWD} directly it exacerbates the symptoms of health conditions and eventually contributes to mortality associated with these them (Hajat, 2017). All metrics were associated with C_{EWD} on their own – specifically, higher t_{10} and t_{min} and smaller t_{sd} were associated with increased risk of C_{EWD}. These effects were diminished in all instances with the inclusion of other confounding variables. Only t_{sd} remained a statistically significant risk factor in the full model. These results did not contradict the findings of the multiple regression model. They added further evidence that those with underlying health conditions, like with the subcategory of participants who are unable to work due to disability or sickness, tend to be associated with narrow thermal variety and increased lower experienced temperature metrics. The study design of this thesis is only able to examine associations. No causal claims can be made from the data alone. However, it is still possible to generate potential causal structures that are consistent with these findings, with a view to generating future programs of research. This will be done below, following an examination of the limitations of the research design.

9.3 Assumptions and limitations

All the regression methods used in this study are linear. The analysis of the statistical validity of the models was given at the end of each of the results chapters, and it was shown that the assumption of linearity held for the data considered. However, most approaches to the relationship between cold exposure and mortality in contemporary epidemiology consider non-linear and lagged exposure effects. The structure of the variables and their collection times means this was not possible or meaningful for this study. However, it would be beneficial to allow for non-linear effects in future designs using experienced temperature. This is discussed further in section 10.2, but it is important to note that the ordering of data collection for this study, where baseline characteristics were collected well in advance of the experienced temperature and activity data, mean contemporary approaches to cold exposure were not possible in this study. Given this limitation, the restriction of the analysis to linear effects is reasonable.

The experienced temperature and thermal variety were summarised into a single variable which accounted for the 5-day study period. The same approach was taken with external temperature. However, as was seen in chapter 8, the time-series temperature data exhibit diurnal variation. Therefore, in averaging across multiple days the diurnal characterises of the time-series are lost. A more complex study design could have used hourly external temperature data, alongside hourly measures of experienced temperature, to understand how they change through the day and week. Again, these are further considered in section 10.2. Unfortunately, since the initial downsampling and processing of the data was a highly time intensive task, more nuanced analysis was not possible in the time frame of the PhD.

A similar consideration applies to the treatment of geographical structure in the data. While this was tested using a multilevel approach and ultimately rejected due to insufficient variance, alternative approaches are possible. Most notably, the findings regarding open solid-fuel fires would have been interesting to test against a measure which estimated the urban or rural characteristics of the household. While it would not have been possible to be more precise than the kilometre rounding of the home location allowed, an approximate categorisation would have been beneficial.

Finally, the primary limitation of the study is the lack of information regarding where the participant was during the study period. The proxy of the activity filtered time-series, i.e. t_{10}^{m} and other metrics, was partially successful but the study program would have benefited from validation in this respect. Whether the participant is indoors, whether at home or elsewhere, is crucial to making policy recommendations and understanding the relation with energy demand more specifically. The present study would have been greatly aided by a variable that assessed whether the participant was indoors or outside. This has been successfully implemented in small scale studies of activity (Kerr et al., 2012) using GPS systems, but this would be unfeasible for a large-scale study such as the present one and could also present privacy issues.

9.4 The overall picture

Across all metrics, including the information provided by the binomial regression, and keeping in mind the limitations and assumptions of the research design, a possible summary model of the findings regarding temperature and health is given in figure 9.1. It incorporates the general consensus from the wider literature, and the findings of this study, to show that as health moves from good to poor the thermal variety of experienced temperature decreases. This is consistent with the finding that the lowest temperatures experienced are less cold as health becomes poor. These findings also hold for what happens as people age or become less active. On top of this narrowing band of experienced temperatures, two regions of harm are imposed. These show that as one moves towards poorer health, the region of potentially harmful temperatures encroaches onto the experienced temperature of the participant. Under this summary model, it is this effect that contributes to the seasonal variation of mortality that is observed in the epidemiological literature.

9.5 Implications

Without appealing to casual mechanisms, it is difficult to generate the possible implications of the findings. Therefore, this section takes a speculative approach as to what the possible underlying causal mechanisms might be which produce the observed research findings. In order to restrict the scope of this discussion to a manageable extent, four primary findings from section 9.1 are focused on.

9.5.1 Age

Cold exposure and thermal variety are shown to decrease with age. Clearly, this association is driven by ageing, which occurs endogenously to the thermal environment of the study participants, as opposed to narrowing thermal variety causing the ageing process. As people age, it is reasonable to assume that the variety of thermal environments they experience on a daily basis reduces. The fact that the data controls for participant activity levels is important. The reduction of thermal variety with age, and the reduced incidence of cold exposure, shows that when activity does occur in older people it is more likely to be environments which are thermally similar to those in which they are sedentary. There is some evidence to suggest that physical activity outdoors confers an added benefit to health over activity indoors (Pasanen et al., 2014). It is plausible that an element of the benefit

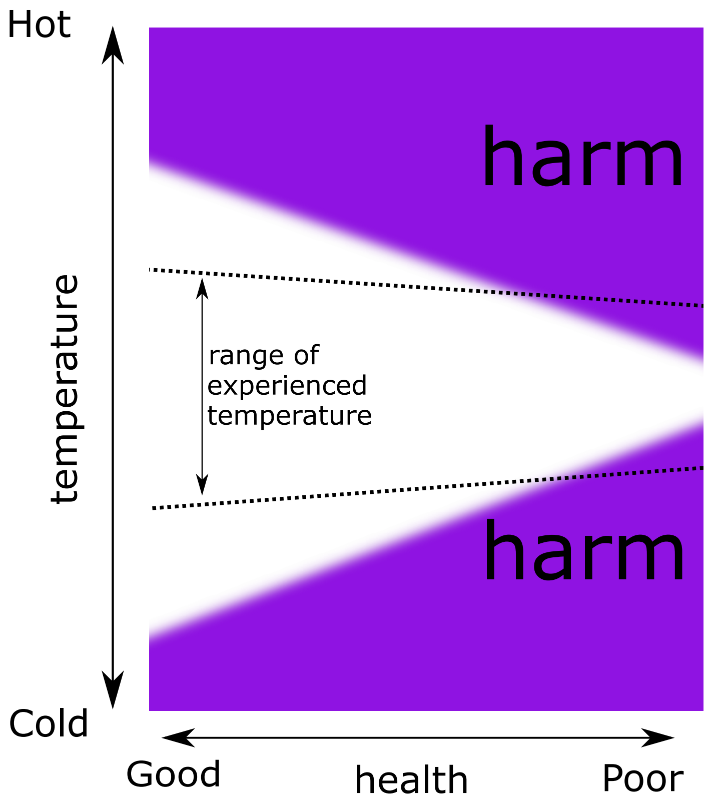


Figure 9.1: A summary of the findings, with information from the literature incorporated. The central dotted lines show that as health moves from good to poor, experienced temperatures narrow. The literature suggests that people in poor health are more vulnerable to cold (and hot) temperatures than people in good health. This is illustrated by the purple regions of harm, which encroach on the range of experienced temperatures as health becomes worse.

of outdoor activity is related to thermal variety, although this hypothesis would require a specific study design.

9.5.2 Activity level

The association between activity level and experienced temperature was very clear across all metrics. Those in the highest quintile of activity had lower experienced temperature and greater thermal variety. This was evident for times at which the participant was sedentary and when active. The most reasonable hypothesis regarding the direction of causality for these variables would be that those who are more active have access to a greater variety of thermal environments, and that they tend to be outside for longer during colder conditions. It is also plausible that the causal direction should go in the other direction i.e. a situation in which someone in a warmer environment would reduce the level of activity. Indeed, the converse has been shown in the literature: Gauthier (2016) found that increased activity is one of the typical responses to being in an increasingly cold environment. However, this has not been found to translate into a positive relationship between BMI and domestic temperatures in the literature, as was discussed in chapter 2.

9.5.3 Health

The effect of health satisfaction was clear for the unfiltered metrics. Diminishing health satisfaction was associated with decreased thermal variety and increased experienced temperature. Thermal variety during sedentary periods also decreased for those who were less happy with their health, and to a lesser extent for the minimum experienced temperature, but the same effect was not seen for the first decile. Those who were unable to work due to sickness or disability showed warmer lower metric temperatures (except for the first decile while sedentary) compared to those who were employed. There was also evidence from the binomial regression that increased risk of having a condition associated with excess winter deaths was associated with diminished thermal variety (but this was not observed for the lower temperature metrics). Again, the most obvious explanation for these findings is that those who are in poor health have less access to thermal varied environments, and avoid or attempted to avoid cold exposure. Any causal mechanisms along the opposite direction, for which low temperature exposure caused health improvement, would likely be of the kind investigated by van Lichtenbelt et al. (2017), in which mild exposure to cold was shown to stimulate metabolic activity. It is certainly plausible that both of these causal directions influence the observed relationships in this study.

9.5.4 Buildings and heating configurations

Finally, the building and heating configurations are considered. Clear results as a function of building type were disappointingly absent from the findings. There was a small effect that showed that those who lived alone experienced lower temperatures and had greater thermal variety than those who live in multiple occupant dwellings. One possible reason for this might be that those who live in homes of multiple occupants have heating patterns that have to cater for more than one person, so cold exposure is less likely. However, this finding requires more detailed investigation to be understood.

The clearest effect came from those who made use of an open solid-fuel fire for heating in winter. For these participants thermal variety was greater, and they had lower minimum and first decile experienced temperatures. The results held for both times at which the participant was sedentary and in general. For this variable it is clear that the housing configuration must influence the experienced temperature, and not the other way around. Some plausible reasons for this have been outlined above, including an absence of the urban heat island effect in more rural communities where open fires are common, poorer building fabric and different home heating configurations for which there is less heating in rooms which are not in regular use. These hypothetical causal pathways require further investigation. Indeed, a validation case study across a variety of homes and sociodemographic types would make an excellent follow-up study. This is further considered in section 10.2. The remainder of this discussion chapter is devoted to considering the wider implications of the model presented above in figure 9.1.

9.5.5 Energy and buildings

The conceptual summary of the thesis findings given in figure 9.1 is useful for considering the potential implications of this study for energy demand reduction in the built environment more generally. The introduction in chapter 1 set out the built environment demand context. Returning briefly to this it is important to recall that in the temperate climates of the Northern Hemisphere, domestic energy demand is dominated by space heating. In the USA, for example there are four times as many heating degree days as cooling degree days EIA (2012). In the EU, domestic space heating accounts for 78% of domestic energy use, at least 60% of which comes directly from fossil fuel sources.

This study has shown how differences in experienced temperature, and thermal variety, are associated with a number of demographic and housing factors. In terms of figure 9.1, reducing harmful exposure necessitates modifying internal temperatures in under–heated dwellings. Furthermore, low thermal variety, especially in winter, may also point to the problem of chronically low experienced temperature. It is vital that attention is focused on at-risk populations who lack the means to avoid harmful cold exposure. Secondary health impacts associated with temperature such mould growth and damp, which are more prevalent in under–heated homes, are also a priority.

In the EU, space cooling is uncommon in homes. Since the majority of domestic energy is expended in home heating this remains the primary target for domestic energy demand reduction. There is the potential to reduce the heating demand temperature of homes with healthy occupants, effectively increasing their thermal variety from below, without increasing the risk to vulnerable individuals. If such individuals could be encouraged to reduce their heating demand temperatures at home, perhaps accompanied by comfort provision using low carbon practices, a significant amount of energy might be saved. From a policy standpoint, such a position is currently controversial given government and health body recommendations typically avoid differentiating between thermal environments for healthy and unhealthy individuals (cf. Jevons et al. (2016)). However, when coupled with the emerging evidence from the thermal comfort literature on the comfort potential of indoor environments which avoid thermal monotony, such a proposal has broader appeal. Indeed, a simple analysis given in the appendix D shows that if the healthiest 25% of the population of Europe could be motivated to reduce their internal domestic temperatures by 1°C, approximately 10 $MtCO_2e$ (around 0.2% of total

EU emissions) could be saved. Heating reduction campaigns could be targeted at healthy, well-off and environmentally conscious portions of the population as a means raising awareness of the climate impacts of CO_2 intensive heating.

Chapter 10

Conclusions

But do you find the change in the seasons affects you without reason?
You've greetings but nothing more to say
And you swear you'd feel much better if only summer'd last forever
But the sky is clear and you're nowhere near and the rain is here again

CATATONIA – DON'T NEED THE SUNSHINE (1998)

This concluding chapter outlines the main outcomes and key findings of the study. Potential avenues for future work are given. The study design as a whole is reflected upon, and the potential wider impact of the work laid out.

10.1 Main conclusions

- The wrist worn AX3 monitor characterises the immediate thermal environment of the wearer. The value it records is governed by a mixture of ambient temperature and heat from the wrist, as well as multiple other factors such as clothing level, local wind speed and incident radiation. In all but a few cases, the reading taken by the AX3 is likely correlated to the thermal sensation of the wearer.

- The AX3 is sufficiently accurate to distinguish differences in experienced temperature between sociodemographic groups and housing and health factors, as well as the average external temperature at the time of wearing.

- The best way to characterise the time-series of temperatures measured by the AX3 in relation to cold exposure is through the lower temperature metrics (t_{10}, t_{min}). The use of metrics screened for high activity (t_{10}^m, t_{min}^m) provides insight as to whether the cold exposure occurs during sedentary periods. These periods likely correspond to times when the participant was indoors.

- The thermal variety (t_{sd}) is also useful as a metric of cold exposure, but it is potentially impacted by the frequency of exposure to warmth.

- The clearest effects are as follows:
  - Cold exposure is measured to decrease with age. Thermal variety also decreases with age.
  - Cold exposure increases with activity level. Thermal variety also increases with activity level.
  - Cold exposure decreases with health satisfaction and whether the participant has conditions associated with excess winter deaths. Thermal variety also decreases with the factors.
  - Cold exposure and thermal variety increase if the household makes use of an open solid fuel fire for heating in winter time. Further investigation is required to understand which factors primarily determine this finding. Possibilities include housing location, fabric and the distribution of heat in bedrooms and other rooms of the home.
  - Cold exposure is greater for those who live alone compared to those who live multiple occupant dwellings.
- Smaller effects are observed for sex differences and they are complex. Men tend to have lower experienced temperatures, but also slightly lower thermal variety.
- Effects related to Body Mass Index and ethnicity are also complex and require further investigation.
- A follow up project of times-series validation is required to associate the specific readings at a given time in a given environment with the activities of daily life of the individual. Many of the experienced temperature times-series vary diurnally with a night time peak, the specific mechanics and the relative importance of wrist temperature and ambient temperature should be understood. Better understanding of the specific effect of clothing level is also important.

10.2 Further work

Several avenues of further work using the processed data became apparent during the research process. One of the primary aims of this thesis was to create a database of experienced temperatures recorded using the AX3. Indeed, the complete database of downsampled and processed data will be returned to the UK Biobank so that other researchers can apply for access to make use of it. The following section highlights a few ideas for further work that could be carried out.

10.2.1 Repeat measure

As it stands the data is based on a single sample of a week for each participant between 2013 and 2015. UK Biobank will soon be releasing data collected in 2018 where participants were asked to wear the device at four instances throughout the year. This would provide insight into the variation of experienced temperature throughout the year within a particular participant. This would provide a step towards being able to make causal statements, when coupled with repeat measures of health satisfaction, for example.

10.2.2 Activity analysis

The team at the University of Newcastle, along with researchers related to the study due to Doherty et al. (2017), have developed an improved processing script which is able to differentiate activity patterns and relate them to activities such as sleeping and exercise. This would provide a more reliable alternative to the method of filtering by median activity level used in this study. Understanding how the experienced temperature relates to specific activities would be extremely valuable.

10.2.3 Rhythmicity

It was evident that certain experienced temperature time-series exhibited greater regularity in their diurnal variation that others. Developing a characterisation of the rhythmicity of the experienced temperature might provide additional insight on top of the metrics of cold exposure and thermal variety, as measured through the simple standard deviation. Certain subjects appear to have repeated thermal routines, much like limit cycles in chaos theory, others have time-series which seem to be affected by noise. What this means for participant health is also of great interest.

10.2.4 Night-shift workers

One excluded portion of the dataset were those who work night-shifts. These participants were excluded because there is evidence that suggests the circadian rhythms are disrupted by night-shift work, which is detectable in the wrist temperature variation. However, during discussions with colleagues, the suggestion was made that this group could be of particular interest from a thermophysiological perspective. A potential follow up study could focus on this group in particular, and attempt to determine whether the experienced temperature disruption plays a part in development of conditions associated with disrupted circadian rhythms.

10.2.5 Showering

In the pilot study using looking at 10-day's worth of data recorded on the AX3 it was evident that showering events could be detected. Data on the frequency of showering could provide insights for two very distinct fields of study. First, hot water heating schedules might be derivable from such data, which could aid in demand side management. Second, there is a link between positive mental health and the maintenance of personal hygiene (Sum et al., 2019). Understanding the frequency of showering in the population, and its relation to depressive episodes and other mental health conditions might prove insightful.

10.2.6 Thermal variety and BMI

The study revealed clear relationships between increased body mass index and reduced thermal variety. Understanding this relationship in greater detail may aid the development of novel treatments for metabolic syndrome and other conditions related to obesity.

10.3 Critical reflection on the research design

Broadly speaking the research program was successful. The data processing and downsampling portion was challenging as it involved the deployment of complex programming techniques. In particular, the set-up of the cluster computing environments was time intensive. However, once the system was set-up it was possible to proceed with the processing using the modified Newcastle University script relatively quickly.

The inclusion of a Pre–analysis plan (PAP) was deemed necessary to guard against the risks of p-hacking. The use of a PAP limits the scope of the tests that can be carried out in the first instance. The limitations it set meant the research proceeded more slowly than otherwise might have happened, but in the end this was valuable as it allowed for further reflection on research questions. It meant that the research proceeded in a more structured way, and overall was a positive addition to the research program. However, it was also necessary that a portion of the study was carried out without strict pre-specification. It is an interesting counter-factual to consider whether the research outcomes might have been different had a PAP not been written. It is hoped that they would have been similar, but the nature of Big Data computing projects is that once the data are available an array of tests can be carried out very quickly. The inclusion of the PAP was therefore most valuable in its effect of reducing the pace of research. This allowed more critical reflection than otherwise might have occurred.

An alternative research design could have focused on using the AX3 in a more controlled case-study environment. This would have provided interesting insight into the relationship between activities of daily life and experienced temperature, as suggested above. However, the research design carried out for this thesis was essential to answering the research questions at the population level.

The consideration of thermal comfort as a research parameter was largely absent from the study. This topic could form the basis of a major follow-up study. The present study would also have benefited from a closer examination of the time variation of experienced temperature. In particular, this could have allowed for sleep periods to be screened. Again, for reasons of scope this was not possible, but would form the basis for future work.

10.4 Thesis outcomes and Impact

The following section gives the principle outcomes of the thesis. Alongside the paper given below, it is expected that two further research papers will be accepted for publication in the coming months. It is hoped that these papers will develop the understanding of themes addressed by this thesis in the literature as a whole. Alongside these papers, the dataset of experienced temperature and activity time-series for 102,342 participants will be available to all UK Biobank researchers.

10.4.1 Research papers

- "Observational evidence of the seasonal and demographic variation in experienced temperature from 77,743 UK Biobank participants" H. R. Kennard, G.M. Huebner , D. Shipworth. J Public Health (Oxf). 2019 Apr 25. fdz025.

10.4.2 Principle conference papers, presentations and posters

- "Thermal Variety and Health". ICEE 2019: International Conference on Environmental Ergonomics 7$^{\text{th}}$ – 12$^{\text{th}}$ July, 2019. Amsterdam, Netherlands.

- "Regression Dilution, Bayesian Analysis and Adaptive Thermal Comfort". Tenth International Windsor Conference : Rethinking Comfort. Windsor, 13$^{\text{th}}$-15$^{\text{th}}$ April 2018. Windsor, UK. (peer-reviewed conference paper)

- "Experienced temperature and health" 2$^{\text{nd}}$ BSA Environment & Health. Conference Poster. 27$^{\text{th}}$ October 2017. University of Cardiff, UK.

- "Experienced temperature and fuel poverty". Fuel Poverty Research Network 4$^{\text{th}}$ Meeting. Lightning Presentation. 6$^{\text{th}}$ & 7$^{\text{th}}$ November 2017. Newcastle upon Tyne, UK.

- "A toolkit for improving the quality and reproducibility of energy research". M. Nicolson, M. Fell, G. Huebner, D. Shipworth, S. Elam, C. Hanmer, H. Kennard. ERSS2017 1$^{\text{st}}$ International Conference on Energy Research & Social Science, 2$^{\text{nd}}$–5$^{\text{th}}$ April 2017, Sitges, Spain.

- "Are we heading towards a replicability crisis in energy efficiency research? A toolkit for improving the quality, transparency and replicability of energy efficiency impact evaluations" G. Huebner, M. Nicolson, M. Fell, D. Shipworth, S. Elam, C. Hanmer, H. Kennard, C. Johnson. In: Proceedings of the European Council for an Energy

Efficient Economy ECEEE 2017 Summer Study on energy efficiency: consumption, efficiency and limits. UKERC: London, UK.

10.4.3 Impacts

Within academia, the impact of this thesis is three-fold. First, from a methods perspective, it has demonstrated that it is possible to measure and analyse the experienced temperature of individuals at a population level. The number of participants involved in this study far exceeds any previous studies of experienced temperature or domestic temperatures in the UK. It is hoped that this thesis has demonstrated the viability of the use of wrist worn temperature sensors in the real world. The method presented here could be of value to health practitioners and researchers who work in related fields such as fuel poverty. Second, the structure of the UK Biobank means that the data outputted by the Big Data processing portion of this thesis will be available to other researchers. This opens up potential avenues for future work using the concept of experienced temperature, some of which were outlined in the previous section. Finally, it contributes to the literature regarding the sociodemographic variation of temperatures in the UK.

Outside academia this study has the potential to contribute practical understanding which could aid the development of commercial heating or thermal comfort control systems. It is conceivable that a wrist worn device could be developed that would automatically regulate heating in order to maximise the efficiency of such systems. As the economy is decarbonised, improvements in the energy efficiency of home heating devices is essential. This is especially true if the UK adopts heating systems based on electric heat pumps, but also applies if bio-gas or hydrogen based systems are preferred.

Beyond this, there is an opportunity to critically engage with policy makers regarding the immediate health impacts of adequate domestic heating for vulnerable populations, and in the longer term, the wider scale impacts of carbon based heating systems for the population as a whole.

Appendix A

Processing

A.1 CWA processing work-flow

Legion set up notes (written by Stuart Grieve of the UCL Research IT Services). The modified Newcastle University CWA processing script is available from `https://github.com/UCL/AX3-temp-output`. This assumes ssh keys are set up correctly and you can scp data to and from research data and Legion.

1. Check out the Newcastle script repository into your Legion home directory:

   ```
   git clone https://github.com/UCL/AX3-temp-output.git
   cd AX3-temp-output
   ```

2. Compile the Java code:

   ```
   module load java/1.8.0_45
   javac java/*.java
   ```

3. Configure a virtual environment. Ensure that the scp command in the script CWA.sh found in the legion folder will connect without password prompts. Note that this is not the job script. An example to do this would be to run these commands while logged on to Legion:

   ```
   echo testing >> testfile.txt
   scp testfile.txt $USER@ssh.rd.ucl.ac.uk:~
   ```

4. Build a list of CWA files to process, with the full path to each file as well as the filename, and one file per line. This file must be stored in the legion folder. The first few lines of this file might look like this:

   ```
   /mnt/gpfs/live/ritd-ag-project-rd00g7-dtshi69/2037069_90001_0_0.CWA
   /mnt/gpfs/live/ritd-ag-project-rd00g7-dtshi69/1346128_90001_0_0.CWA
   /mnt/gpfs/live/ritd-ag-project-rd00g7-dtshi69/1687108_90001_0_0.CWA
   /mnt/gpfs/live/ritd-ag-project-rd00g7-dtshi69/1346135_90001_0_0.CWA
   /mnt/gpfs/live/ritd-ag-project-rd00g7-dtshi69/1687201_90001_0_0.CWA
   ```

5. Run the template script to create the job script. The template script, build_job_script.py takes 3 arguments, your username and the path and filename of the list of files created in step 5. In this example I am using the username testuser and a file list called filelist_apr14oct14.txt, which is found in the legion folder. This script needs to be run inside the virtual environment created in step 3.

```
module load python2/recommended
source ~/venv/bin/activate
python build_job_script.py testuser ~/AX3-temp-output/legion/
    filelist_apr14oct14.txt
```

6. Change any other parameters in the job script that need changed, e.g. the size of the job array, the wallclock time, or the memory allocation.

7. Submit the job as normal:

```
qsub legion/CWA.job
```

8. Its progress can be monitored using watch:

```
watch -n1 --differences qstat
```

9. When a job successfully completes, the results will be stored in `/mnt/gpfs/live/ritd-ag-project-rd00g7-dtshi69/results/` inside a folder called the name of the input CWA file. The CWA file that has been successfully processed is also moved into the folder `/mnt/gpfs/live/ritd-ag-project-rd00g7-dtshi69/done/`. If a job fails for any reason its log files can be seen using cat on legion:

```
cat ~/Scratch/output/CWA.e<job id>.<task id>
cat ~/Scratch/output/CWA.o<job id>.<task id>
```

   where <job id> is the job id created when the job was qsubbed, and <task id> is the array job number corresponding to the individual file being processed.

Appendix B

Predictors	*Subcategory*	t_{10}	t^m_{10}	t^f_{10}
Intercept		23.62 [23.45 – 23.80]	26.51 [26.35 – 26.67]	26.67 [26.51 – 26.83]
External temperature C		0.12 [0.11 – 0.12]	0.08 [0.08 – 0.08]	0.09 [0.08 – 0.09]
Age (40-49)	50-59	0.03 [0.03 – 0.03]	0.02 [0.02 – 0.02]	0.01 [0.01 – 0.02]
Sex (Female)	Male	-0.02 [-0.05 – 0.00]	-0.08 [-0.11 – -0.05]	-0.11 [-0.14 – -0.08]
Ethnic background (White)	Mixed	-0.11 [-0.30 – 0.09]	-0.11 [-0.29 – 0.07]	-0.08 [-0.26 – 0.10]
	Asian	0.30 [0.15 – 0.45]	0.15 [0.01 – 0.30]	0.15 [0.01 – 0.29]
	Black	0.10 [-0.07 – 0.26]	0.12 [-0.03 – 0.27]	0.20 [0.05 – 0.36]
	Chinese	-0.51 [-0.82 – -0.19]	-0.66 [-0.95 – -0.37]	-0.56 [-0.85 – -0.27]
	Other ethnic group	-0.05 [-0.25 – 0.14]	-0.13 [-0.32 – 0.05]	-0.10 [-0.28 – 0.08]
	Do not know	-0.04 [-0.91 – 0.83]	-0.21 [-1.02 – 0.59]	-0.25 [-1.05 – 0.56]
	Prefer not to answer	0.10 [-0.19 – 0.38]	-0.01 [-0.27 – 0.25]	-0.04 [-0.30 – 0.22]
Household Income (less than 18,000)	18,000 to 30,999	0.04 [-0.01 – 0.09]	0.04 [-0.00 – 0.09]	0.06 [0.01 – 0.11]
	31,000 to 51,999	-0.01 [-0.07 – 0.04]	0.00 [-0.05 – 0.05]	0.02 [-0.03 – 0.06]
	52,000 to 100,000	0.00 [-0.06 – 0.05]	0.00 [-0.05 – 0.05]	0.01 [-0.04 – 0.06]
	Greater than 100,000	0.01 [-0.06 – 0.09]	0.00 [-0.07 – 0.06]	0.01 [-0.06 – 0.07]
	Prefer not to say	0.02 [-0.05 – 0.08]	0.06 [0.00 – 0.13]	0.07 [0.01 – 0.13]
	Do not know	0.14 [0.04 – 0.23]	0.13 [0.04 – 0.22]	0.16 [0.07 – 0.25]
Tenure type (Own outright)	None of above	0.16 [-0.07 – 0.40]	0.03 [-0.19 – 0.25]	0.04 [-0.18 – 0.26]
	Prefer not to answer	-0.12 [-0.39 – 0.15]	-0.19 [-0.44 – 0.06]	-0.17 [-0.42 – 0.08]
	Mortgage	0.12 [0.09 – 0.16]	0.04 [0.00 – 0.07]	0.04 [0.01 – 0.07]
	Rent Local Authority	0.36 [0.26 – 0.45]	0.15 [0.06 – 0.24]	0.16 [0.07 – 0.24]
	Rent private	0.05 [-0.06 – 0.15]	-0.08 [-0.17 – 0.02]	-0.06 [-0.15 – 0.04]
	Shared	0.14 [-0.15 – 0.44]	0.01 [-0.26 – 0.29]	0.01 [-0.27 – 0.28]
	Rent free	0.27 [0.09 – 0.45]	0.05 [-0.12 – 0.22]	0.05 [-0.12 – 0.22]
Accommodation Type (House or Bungalow)	None of above	-0.20 [-0.63 – 0.22]	-0.14 [-0.53 – 0.25]	-0.13 [-0.52 – 0.27]
	Prefer not to answer	-0.09 [-1.20 – 1.02]	-0.22 [-1.24 – 0.81]	-0.35 [-1.38 – 0.68]
	Flat	0.22 [0.16 – 0.28]	0.14 [0.09 – 0.19]	0.13 [0.07 – 0.18]
	Temporary	-0.04 [-0.56 – 0.49]	0.08 [-0.41 – 0.57]	0.00 [-0.49 – 0.49]
Household Size (one)	Two	0.11 [0.07 – 0.15]	0.19 [0.15 – 0.23]	0.18 [0.14 – 0.22]
	Three	0.08 [0.03 – 0.14]	0.14 [0.09 – 0.18]	0.15 [0.10 – 0.20]
	4+	0.04 [-0.01 – 0.10]	0.10 [0.05 – 0.15]	0.12 [0.07 – 0.17]
Employment Type (Employed)	None of above	-0.19 [-0.40 – 0.02]	-0.20 [-0.39 – -0.00]	-0.23 [-0.43 – -0.04]
	Prefer not to answer	0.13 [-0.27 – 0.52]	0.17 [-0.19 – 0.54]	0.19 [-0.17 – 0.55]
	Retired	-0.03 [-0.07 – 0.01]	0.02 [-0.01 – 0.06]	-0.01 [-0.04 – 0.03]
	Looking after home...	-0.10 [-0.17 – -0.02]	-0.06 [-0.13 – 0.01]	-0.04 [-0.11 – 0.02]
	Unable to work...	0.50 [0.39 – 0.60]	0.17 [0.07 – 0.27]	0.09 [-0.01 – 0.19]
	Unemployed	0.04 [-0.09 – 0.17]	-0.07 [-0.19 – 0.05]	-0.10 [-0.22 – 0.02]
	Doing unpaid or voluntary	-0.08 [-0.15 – -0.01]	-0.04 [-0.11 – 0.02]	-0.04 [-0.11 – 0.02]
	Full or part-times student	-0.09 [-0.23 – 0.05]	-0.11 [-0.24 – 0.03]	-0.11 [-0.24 – 0.02]
Fuel type (Gas hob or gas cooker)	Gas fire	-0.02 [-0.07 – 0.03]	0.00 [-0.05 – 0.05]	0.00 [-0.05 – 0.05]
	An open solid fuel fire	-0.37 [-0.46 – -0.29]	-0.18 [-0.26 – -0.10]	-0.14 [-0.22 – -0.06]
	Gas Hob & Gas Fire	-0.02 [-0.05 – 0.02]	0.00 [-0.03 – 0.03]	0.00 [-0.04 – 0.03]
	Gas Hob & Open s.f. fire	-0.25 [-0.31 – -0.18]	-0.10 [-0.16 – -0.05]	-0.08 [-0.14 – -0.02]
	Gas Fire & Open s.f. fire	-0.57 [-0.85 – -0.29]	-0.43 [-0.69 – -0.17]	-0.39 [-0.65 – -0.13]
	Hob & Gas Fire & O.s.f. fire	-0.28 [-0.41 – -0.15]	-0.23 [-0.35 – -0.11]	-0.22 [-0.34 – -0.10]
	None of the above	0.07 [0.03 – 0.11]	0.06 [0.03 – 0.10]	0.06 [0.02 – 0.10]
	Prefer not to answer	0.47 [-0.18 – 1.12]	0.38 [-0.23 – 0.98]	0.40 [-0.20 – 1.00]
	Do not know	-0.39 [-1.47 – 0.69]	-0.67 [-1.67 – 0.33]	-0.70 [-1.70 – 0.30]

Figure B.1: The complete output for the PAP variables for the metrics first decile of the experienced temperature, t_{10}, t^m_{10} and t^f_{10}, which correspond to different activity exclusion criteria. Significance levels: light green $p < 0.01$, mid green $p < 0.001$, dark green $p < 1 \times 10^{-9}$.

Predictors	*Subcategory*	t_{min}	t^m_{min}	t_{10}	t^m_{10}	t_μ	t^m_μ	t_{max}	t^m_{max}
Intercept		18.95	21.15	26.00	27.65	30.31	32.02	36.94	36.74
External temperature °C		0.25	0.19	0.12	0.08	0.05	0.01	0.06	0.03
Age (40-49)	50-59	0.37	0.32	0.06	0.06	0.01	-0.01	-0.11	-0.09
	60-69	0.77	0.71	0.16	0.18	0.06	0.04	-0.19	-0.12
	70-79	1.17	1.05	0.33	0.28	0.13	0.11	-0.18	-0.09
Sex (Female)	Male	-0.61	-0.28	-0.10	-0.10	-0.17	-0.31	-0.21	-0.24
Ethnic background (White)	Mixed	-0.03	-0.04	-0.08	-0.11	0.01	0.05	0.00	-0.02
	Asian	0.89	0.43	0.23	0.12	0.27	0.28	0.20	-0.03
	Black	0.35	0.06	0.09	0.11	0.28	0.34	0.03	-0.02
	Chinese	-0.24	-0.33	-0.36	-0.62	-0.24	-0.24	-0.37	-0.38
	Other ethnic group	0.33	-0.02	0.00	-0.11	0.04	0.05	0.02	-0.07
	Do not know	-0.81	-0.51	0.02	-0.20	-0.10	-0.09	-0.13	-0.25
	Prefer not to answer	0.03	-0.15	0.06	-0.02	0.00	-0.06	0.01	-0.02
Household Income (less than 18,000)	18,000 to 30,999	0.00	0.04	0.06	0.05	0.03	0.01	0.08	0.06
	31,000 to 51,999	-0.12	-0.07	0.03	0.01	0.00	-0.01	0.08	0.06
	52,000 to 100,000	-0.26	-0.14	0.03	0.01	0.01	-0.02	0.11	0.07
	Greater than 100,000	-0.46	-0.21	0.06	0.00	0.02	-0.04	0.19	0.12
	Prefer not to say	-0.02	-0.02	0.05	0.08	0.04	0.04	0.08	0.08
	Do not know	0.28	0.24	0.19	0.15	0.11	0.06	-0.01	0.00
Tenure type (Own outright)	None of above	0.24	0.15	0.09	0.00	0.02	-0.04	-0.01	-0.04
	Prefer not to answer	-0.16	-0.25	-0.14	-0.19	-0.16	-0.16	-0.31	-0.23
	Mortgage	0.07	0.03	0.06	0.01	-0.02	-0.06	-0.07	-0.05
	Rent Local Authority	0.43	0.27	0.23	0.12	0.03	-0.09	-0.16	-0.16
	Rent private	0.01	-0.08	-0.03	-0.10	-0.08	-0.14	-0.12	-0.08
	Shared	0.05	0.06	-0.01	-0.03	-0.09	-0.10	-0.23	-0.29
	Rent free	0.31	0.12	0.16	0.01	0.03	-0.04	-0.10	-0.07
Accommodation Type (House or Bungalow)	None of above	-0.14	0.03	-0.19	-0.14	-0.21	-0.25	-0.48	-0.31
	Prefer not to answer	0.21	0.33	-0.52	-0.36	-0.58	-0.74	-0.45	-0.58
	Flat	0.10	0.10	0.19	0.13	0.09	0.03	0.01	0.02
	Temporary	-0.22	-0.14	-0.07	0.07	-0.08	-0.08	0.01	-0.20
Household Size (one)	Two	0.14	0.18	0.11	0.19	0.07	0.08	-0.03	-0.04
	Three	0.11	0.14	0.09	0.14	0.02	0.03	-0.06	-0.06
	4+	-0.05	0.07	0.07	0.10	0.01	0.03	-0.07	-0.06
Employment Type (Employed)	None of above	-0.12	-0.17	-0.2	-0.2	-0.17	-0.14	-0.17	-0.12
	Prefer not to answer	0.30	0.55	0.17	0.19	0.24	0.29	0.18	0.09
	Retired	-0.04	0.06	-0.05	0.03	-0.02	0.00	-0.05	-0.03
	Looking after home and/or family	0.09	-0.01	-0.04	-0.04	-0.04	-0.01	-0.03	0.03
	Unable to work due to sickness/disability	0.60	0.42	0.21	0.1	0.06	-0.01	-0.01	-0.02
	Unemployed	0.13	0.13	-0.02	-0.09	-0.07	-0.11	-0.09	-0.04
	Doing unpaid or voluntary	-0.07	-0.11	-0.07	-0.03	-0.02	0.02	-0.05	-0.01
	Full or part-time student	-0.27	-0.26	-0.11	-0.11	-0.06	-0.05	-0.05	0.01

Table B.1: The regression results for multiple experienced temperature metrics. Significance levels: light blue $p < 0.01$, mid blue $p < 0.001$, dark blue $p < 1 \times 10^{-9}$. N=77,762.

Predictors	*Subcategory*	t_{min}	t^m_{min}	t_{10}	t^m_{10}	t_μ	t^m_μ	t_{max}	t^m_{max}
Fuel type (Gas hob or Gas cooker)	Gas fire	0.08	0.03	-0.02	0.00	0.01	0.01	0.10	0.04
	An open solid fuel fire	-0.31	-0.40	-0.27	-0.15	-0.09	-0.01	0.20	0.18
	Gas Hob & Gas Fire	0.00	0.00	-0.02	0.00	0.00	0.00	0.10	0.06
	Gas Hob & Open solid fuel fire	-0.23	-0.20	-0.18	-0.08	-0.04	0.01	0.19	0.19
	Gas Fire & Open solid fuel fire	-0.33	-0.53	-0.45	-0.39	-0.16	-0.02	0.28	0.28
	Hob & Gas Fire & Open solid fuel fire	-0.27	-0.36	-0.23	-0.22	-0.11	-0.05	0.18	0.18
	None of the above	0.02	0.03	0.06	0.06	0.04	0.02	0.03	0.01
	Prefer not to answer	0.53	0.65	0.50	0.40	0.18	0.04	0.27	0.14
	Do not know	0.55	0.10	-0.47	-0.72	-0.70	-0.46	-1.23	-0.84
BMI (normal)	Underweight	0.07	0.00	-0.11	-0.17	-0.01	0.05	0.15	0.13
	Overweight	0.27	0.18	0.15	0.05	-0.07	-0.15	-0.10	-0.14
	Obese	0.49	0.21	0.19	-0.17	-0.30	-0.41	-0.47	-0.49
Activity level (0-20)	20-40	-0.76	-0.50	-0.40	-0.07	-0.21	0.00	0.08	0.04
	40-60	-1.17	-0.84	-0.69	-0.16	-0.36	-0.02	0.11	0.07
	60-80	-1.64	-1.10	-0.97	-0.26	-0.53	-0.06	0.14	0.11
	80-100	-2.46	-1.60	-1.44	-0.45	-0.78	-0.10	0.20	0.17
	Observations	77762	77762	77762	77762	77762	77762	77762	77762
	R^2	0.21	0.15	0.14	0.06	0.05	0.02	0.02	0.02

Table B.2: Continuation of figure B.1. Significance levels: light blue $p < 0.01$, mid blue $p < 0.001$, dark blue $p < 1 \times 10^{-9}$. N=77,762.

Predictors	*Subcategory*	t_{min}	t^m_{min}	t_{10}	t^m_{10}	t_μ	t^m_μ	t_{max}	t^m_{max}
Health satisfaction	Very happy	0.17	0.08	0.06	0.01	0.01	-0.01	-0.08	-0.05
	Moderately happy	0.37	0.20	0.14	0.03	0.01	-0.05	-0.11	-0.08
	Moderately unhappy	0.49	0.26	0.20	0.05	-0.01	-0.11	-0.22	-0.17
	Very unhappy	0.59	0.29	0.27	0.07	0.03	-0.11	-0.07	-0.03
	Extremely unhappy	0.73	0.38	0.36	0.14	0.14	0.03	0.03	0.00
	Prefer not to answer	-1.08	-0.50	-0.39	-0.21	-0.27	-0.17	-0.53	-0.49
	Do not know	0.10	0.18	0.20	0.14	0.11	0.09	0.00	-0.01
Financial Situation satisfaction (Extremely happy)	Very happy	0.01	0.01	0.02	0.01	0.03	0.04	0.04	0.02
	Moderately happy	0.02	-0.02	0.01	0.00	0.02	0.02	0.03	0.03
	Moderately unhappy	0.05	-0.02	0.05	0.01	0.02	-0.01	0.00	0.01
	Very unhappy	0.23	0.09	0.20	0.13	0.10	0.04	0.00	0.00
	Extremely unhappy	0.18	0.18	0.10	0.03	0.02	-0.05	0.07	-0.01
	Prefer not to answer	-0.25	-0.15	0.03	-0.26	-0.10	-0.22	-0.32	-0.20
	Do not know	-0.20	0.11	0.20	0.08	0.01	-0.12	-0.36	-0.27
Heating type (Gas central heating)	Electric storage heaters	0.06	0.03	0.06	0.10	0.08	0.08	-0.17	-0.11
	Oil (kerosene) central heating	-0.20	-0.22	-0.14	-0.06	-0.01	0.04	0.07	0.08
	Portable gas or paraffin heaters	0.18	0.35	-0.05	0.12	0.03	0.25	-0.17	0.06
	Solid fuel central heating	-0.60	-0.65	-0.43	-0.14	-0.23	-0.19	-0.08	-0.16
	Open fire without central heating	-0.22	-0.15	-0.18	-0.23	-0.19	-0.23	-0.26	-0.27
	Three heating types	1.09	1.36	0.49	0.15	0.31	0.28	-0.67	-0.37
	None of the above	-0.12	-0.17	-0.07	-0.07	-0.04	-0.06	-0.01	0.01
	Prefer not to answer	0.55	0.50	0.78	0.70	0.50	0.44	0.76	0.08
	Do not know	-0.67	0.13	-0.37	-0.55	-0.57	-0.68	-0.16	-0.45
	Observations	37730	37730	37730	37730	37730	37730	37730	37730
	Adjusted R^2	0.21	0.14	0.14	0.06	0.05	0.02	0.02	0.02

Table B.3: The regression results for the additional variables across different experienced temperature metrics. Significance levels: light blue $p < 0.01$, mid blue $p < 0.001$, dark blue $p < 1 \times 10^{-9}$. N=37,730.

Predictors	*Subcategory*	t_{sd}	t^m_{sd}	t_{range}	t^m_{range}	t_{iqr}	t^m_{iqr}
Intercept		3.46	3.10	18.00	15.59	5.45	4.88
External temperature °C		-0.05	-0.04	-0.19	-0.16	-0.10	-0.08
Age (40-49)	50-59	-0.06	-0.06	-0.49	-0.42	-0.06	-0.08
	60-69	-0.10	-0.12	-0.95	-0.83	-0.08	-0.15
	70-79	-0.16	-0.15	-1.34	-1.15	-0.14	-0.15
Sex (Female)	Male	-0.05	-0.09	0.40	0.04	-0.30	-0.23
Ethnic background (White)	Mixed	0.07	0.07	0.03	0.02	0.18	0.17
	Asian	-0.01	0.03	-0.69	-0.47	0.20	0.06
	Black	0.09	0.09	-0.32	-0.08	0.28	0.13
	Chinese	0.11	0.20	-0.12	-0.05	0.26	0.30
	Other ethnic group	0.03	0.09	-0.31	-0.06	0.15	0.17
	Do not know	0.08	0.11	0.67	0.26	0.05	0.37
	Prefer not to answer	-0.05	-0.01	-0.02	0.12	-0.17	-0.03
Household Income (less than 18,000)	18,000 to 30,999	-0.02	-0.02	0.08	0.02	-0.07	-0.05
	31,000 to 51,999	-0.01	-0.01	0.20	0.13	-0.07	-0.03
	52,000 to 100,000	-0.01	-0.01	0.37	0.21	-0.11	-0.04
	Greater than 100,000	-0.02	-0.01	0.66	0.33	-0.19	-0.09
	Prefer not to say	-0.01	-0.02	0.10	0.09	-0.04	-0.07
	Do not know	-0.07	-0.06	-0.29	-0.24	-0.14	-0.11
Tenure type (Own outright)	None of above	-0.07	-0.02	-0.25	-0.19	-0.15	-0.06
	Prefer not to answer	-0.01	0.01	-0.15	0.02	0.03	0.05
	Mortgage	-0.05	-0.03	-0.14	-0.09	-0.14	-0.06
	Rent Local Authority	-0.16	-0.11	-0.60	-0.43	-0.33	-0.23
	Rent private	-0.04	0.00	-0.13	0.00	-0.09	0.00
	Shared	-0.07	-0.06	-0.28	-0.35	-0.17	-0.17
	Rent free	-0.09	-0.02	-0.41	-0.19	-0.16	-0.05
Accommodation Type (House or Bungalow)	None of above	-0.05	-0.05	-0.34	-0.34	-0.17	-0.1
	Prefer not to answer	-0.17	-0.22	-0.66	-0.91	-0.43	-0.51
	Flat	-0.07	-0.05	-0.08	-0.07	-0.18	-0.10
	Temporary	0.02	-0.03	0.23	-0.06	0.00	-0.10
Household Size (one)	Two	-0.04	-0.07	-0.17	-0.22	-0.07	-0.11
	Three	-0.05	-0.06	-0.18	-0.19	-0.09	-0.10
	4+	-0.03	-0.04	-0.03	-0.13	-0.06	-0.07
Employment Type (Employed)	None of above	0.03	0.02	-0.05	0.05	0.11	0.03
	Prefer not to answer	0.03	-0.01	-0.13	-0.46	0.20	0.01
	Retired	0.01	-0.02	0.00	-0.09	0.02	-0.04
	Looking after home and/or family	0.02	0.02	-0.11	0.04	0.08	0.05
	Unable to work due to sickness/disability	-0.10	-0.05	-0.61	-0.44	-0.17	-0.07
	Unemployed	-0.02	0.00	-0.23	-0.17	-0.03	0.02
	Doing unpaid or voluntary	0.04	0.03	0.01	0.10	0.08	0.07
	Full or part-time student	0.04	0.04	0.23	0.27	0.08	0.03

Table B.4: The regression results for multiple thermal variety metrics. Significance levels: light red $p < 0.01$, mid red $p < 0.001$, dark red $p < 1 \times 10^{-9}$. N=77,762.

Predictors	*Subcategory*	t_{sd}	$t^{m}{}_{sd}$	t_{range}	$t^{m}{}_{range}$	t_{iqr}	$t^{m}{}_{iqr}$
Fuel type (Gas hob or Gas cooker)	Gas fire	0.01	0.01	0.02	0.01	0.05	0.00
	An open solid fuel fire	0.12	0.07	0.51	0.58	0.24	0.08
	Gas Hob & Gas Fire	0.01	0.01	0.09	0.07	0.02	0.00
	Gas Hob & Open solid fuel fire	0.09	0.06	0.41	0.39	0.18	0.09
	Gas Fire & Open solid fuel fire	0.21	0.19	0.62	0.81	0.40	0.32
	Hob & Gas Fire & Open solid fuel fire	0.08	0.09	0.45	0.54	0.14	0.15
	None of the above	-0.01	-0.01	0.01	-0.02	-0.01	-0.02
	Prefer not to answer	-0.21	-0.2	-0.26	-0.51	-0.38	-0.47
	Do not know	-0.18	-0.15	-1.78	-0.94	-0.13	-0.32
BMI (normal)	Underweight	0.11	0.11	0.07	0.13	0.39	0.22
	Overweight	-0.18	-0.14	-0.38	-0.32	-0.50	-0.30
	Obese	-0.37	-0.24	-0.96	-0.70	-0.94	-0.53
Activity level percentile (0-20)	20-40	0.14	0.04	0.84	0.54	0.19	0.05
	40-60	0.24	0.08	1.28	0.91	0.32	0.09
	60-80	0.33	0.11	1.78	1.20	0.45	0.14
	80-100	0.50	0.19	2.66	1.77	0.69	0.25
	Observations	77762	77762	77762	77762	77762	77762
	R^2 / adjusted R^2	0.24	0.18	0.17	0.13	0.20	0.15

Table B.5: Continuation of table B.4. Significance levels: light red $p < 0.01$, mid red $p < 0.001$, dark red $p < 1 \times 10^{-9}$. N=77,762.

Predictors	*Subcategory*	t_{sd}	$t^{m}{}_{sd}$	t_{range}	$t^{m}{}_{range}$	t_{iqr}	$t^{m}{}_{iqr}$
Health satisfaction	Very happy	-0.04	-0.02	-0.25	-0.13	-0.05	-0.03
	Moderately happy	-0.10	-0.05	-0.48	-0.28	-0.14	-0.09
	Moderately unhappy	-0.15	-0.09	-0.71	-0.43	-0.26	-0.16
	Very unhappy	-0.16	-0.08	-0.66	-0.33	-0.32	-0.10
	Extremely unhappy	-0.15	-0.07	-0.70	-0.38	-0.20	-0.13
	Prefer not to answer	0.05	-0.05	0.55	0.01	0.36	-0.22
	Do not know	-0.04	-0.03	-0.09	-0.19	-0.06	-0.05
Financial Situation satisfaction (Extremely happy)	Very happy	0.01	0.02	0.03	0.00	0.04	0.05
	Moderately happy	0.01	0.02	0.00	0.06	0.00	0.05
	Moderately unhappy	-0.02	0.01	-0.04	0.03	-0.07	0.03
	Very unhappy	-0.07	-0.03	-0.23	-0.09	-0.17	-0.05
	Extremely unhappy	-0.06	-0.03	-0.11	-0.19	-0.18	-0.03
	Prefer not to answer	-0.09	-0.02	-0.07	-0.04	-0.16	-0.17
	Do not know	-0.12	-0.09	-0.16	-0.38	-0.32	-0.18
Heating type (Gas central heating)	Electric storage heaters	-0.01	-0.02	-0.23	-0.14	0.01	-0.06
	Oil (kerosene) central heating	0.09	0.07	0.27	0.30	0.17	0.09
	Portable gas or paraffin heaters	0.17	0.04	-0.34	-0.29	0.27	0.04
	Solid fuel central heating	0.09	-0.01	0.52	0.49	0.09	-0.07
	open fire without central heating	-0.02	-0.01	-0.04	-0.12	-0.03	0.02
	Three heating types	-0.17	-0.05	-1.77	-1.74	-0.25	-0.14
	None of the above	-0.01	0.00	0.11	0.18	-0.05	-0.07
	Prefer not to answer	-0.19	-0.22	0.20	-0.41	-0.25	-0.29
	Do not know	-0.17	-0.19	0.51	-0.58	-0.13	-0.24
	Observations	37730	37730	37730	37730	37730	37730
	Adjusted R^2	0.24	0.19	0.17	0.13	0.20	0.15

Table B.6: The regression results for the additional variables across different thermal variety metrics. Significance levels: light red $p < 0.01$, mid red $p < 0.001$, dark red $p < 1 \times 10^{-9}$. N=37,730.

Variable: Subcategory	$d(t_{min})$	$d(t^m_{min})$	$d(t_{10})$	$d(t^m_{10})$
External temperature °C	0.08 [0.07 – 0.08]	0.06 [0.06 – 0.07]	0.06 [0.06 – 0.06]	0.04 [0.04 – 0.05]
Age: 50-59	0.11 [0.09 – 0.14]	0.11 [0.08 – 0.14]	0.03 [0.00 – 0.06]	0.03 [0.00 – 0.06]
Age: 60-69	0.23 [0.20 – 0.26]	0.24 [0.21 – 0.27]	0.08 [0.05 – 0.11]	0.10 [0.07 – 0.13]
Age: 70-79	0.35 [0.32 – 0.38]	0.36 [0.33 – 0.39]	0.16 [0.13 – 0.19]	0.15 [0.11 – 0.18]
Sex: Male	-0.18 [-0.20 – -0.17]	-0.09 [-0.11 – -0.08]	-0.05 [-0.06 – -0.04]	-0.05 [-0.07 – -0.04]
Ethnicity: Asian	0.27 [0.20 – 0.33]	0.15 [0.08 – 0.22]	0.11 [0.04 – 0.18]	0.07[-0.01 – 0.14]
Ethnicity: Black	0.10 [0.03 – 0.18]	0.02 [-0.05 – 0.10]	0.04 [-0.03 – 0.12]	0.06[-0.02 – 0.14]
Ethnicity: Chinese	-0.07[-0.21 – 0.07]	-0.11 [-0.26 – 0.03]	-0.17[-0.32 – -0.03]	-0.33 [-0.48 – -0.17]
Income: 31,000 to 51,999	-0.04 [-0.06 – -0.01]	-0.02 [-0.05 – 0.00]	0.01[-0.01 – 0.04]	0.00[-0.02 – 0.03]
Income: 52,000 to 100,000	-0.08 [-0.10 – -0.05]	-0.05 [-0.07 – -0.02]	0.02[-0.01 – 0.04]	0.01[-0.02 – 0.03]
Income: Greater than 100,000	-0.14 [-0.17 – -0.10]	-0.07 [-0.11 – -0.04]	0.03[-0.01 – 0.06]	0.00[-0.03 – 0.04]
Income: Do not know	0.08 [0.04 – 0.13]	0.08 [0.04 – 0.13]	0.09 [0.05 – 0.14]	0.08 [0.03 – 0.12]
Tenure: Rent Local Authority	0.13 [0.09 – 0.17]	0.09 [0.05 – 0.14]	0.11 [0.07 – 0.15]	0.06 [0.02 – 0.11]
Accommodation: Flat	0.03[0.00 – 0.05]	0.03 [0.01 – 0.06]	0.09 [0.06 – 0.12]	0.07 [0.04 – 0.10]
Household size: two	0.04 [0.02 – 0.06]	0.06 [0.04 – 0.08]	0.05 [0.03 – 0.07]	0.10 [0.08 – 0.12]
Household size: three	0.03 [0.01 – 0.06]	0.05 [0.02 – 0.07]	0.04 [0.02 – 0.07]	0.07 [0.05 – 0.10]
Household size: 4+	-0.01 [-0.04 – 0.01]	0.02 [-0.00 – 0.05]	0.03 [0.01 – 0.06]	0.06 [0.03 – 0.08]
Employment: Unable to work because of sickness/ disability	0.18 [0.13 – 0.23]	0.14 [0.09 – 0.19]	0.10 [0.05 – 0.15]	0.05 [0.00 – 0.10]
Fuel type: An open solid fuel fire	-0.09 [-0.13 – -0.05]	-0.14 [-0.18 – -0.10]	-0.13[-0.17 – -0.09]	-0.08 [-0.12 – -0.04]
Fuel type: Hob & Open fire	-0.07 [-0.10 – -0.04]	-0.07 [-0.10 – -0.04]	-0.09 [-0.12 – -0.06]	-0.04 [-0.07 – -0.01]
Fuel type: Gas Fire & Open fire	-0.10 [-0.22 – 0.03]	-0.18 [-0.31 – -0.05]	-0.22 [-0.35 – -0.08]	-0.21 [-0.35 – -0.07]
Fuel type: Hob & Gas Fire & Open fire	-0.08 [-0.14 – -0.02]	-0.12 [-0.18 – -0.06]	-0.11 [-0.17 – -0.05]	-0.11 [-0.18 – -0.05]
Fuel type: None of the above	0.01 [-0.01 – 0.02]	0.01 [-0.01 – 0.03]	0.03 [0.01 – 0.05]	0.03 [0.01 – 0.05]
BMI: overweight	0.08 [0.07 – 0.09]	0.06 [0.05 – 0.08]	0.07 [0.06 – 0.09]	0.03 [0.01 – 0.04]
BMI: obese	0.15 [0.09 – 0.20]	0.07 [0.01 – 0.13]	0.09 [0.03 – 0.15]	-0.09 [-0.15 – -0.03]
Activity: 20-40	-0.23 [-0.25 – -0.21]	-0.17 [-0.19 – -0.15]	-0.19 [-0.21 – -0.17]	-0.04 [-0.06 – -0.01]
Activity: 40-60	-0.35 [-0.37 – -0.33]	-0.29 [-0.31 – -0.26]	-0.33 [-0.35 – -0.31]	-0.08 [-0.11 – -0.06]
Activity: 60-80	-0.49 [-0.51 – -0.47]	-0.37 [-0.39 – -0.35]	-0.47 [-0.49 – -0.45]	-0.14 [-0.16 – -0.12]
Activity: 80-100	-0.73 [-0.75 – -0.71]	-0.54 [-0.57 – -0.52]	-0.69 [-0.72 – -0.67]	-0.24 [-0.26 – -0.22]
Health: Moderately happy	0.11 [0.07 – 0.15]	0.07 [0.02 – 0.11]	0.07 [0.02 – 0.11	0.02 [-0.03 – 0.06]
Health: Moderately unhappy	0.15 [0.09 – 0.20]	0.09 [0.04 – 0.14]	0.09 [0.04 – 0.15]	0.03 [-0.03 – 0.08]
Health: Very unhappy	0.18 [0.10 – 0.26]	0.10 [0.02 – 0.18]	0.13 [0.05 – 0.22]	0.04 [-0.05 – 0.13]
Health: Extremely unhappy	0.22 [0.10 – 0.34]	0.13 [0.00 – 0.26]	0.17 [0.05 – 0.30]	0.08 [-0.06 – 0.21]
Heating: Solid fuel central heating	-0.18 [-0.34 – -0.02]	-0.22[-0.38 – -0.06]	-0.21 [-0.37 – -0.04]	-0.07[-0.24 – 0.10]

Table B.7: Effect size estimates for t_{min}, t^m_{min}, t_{10} and t^m_{10}, for significant subcategories only. Significance levels denoted by colour: light yellow $p < 0.01$, mid yellow $p < 0.001$, deep yellow $p < 1 \times 10^{-9}$. N=77,762 for variables above the dark line, and N=37,730 below the dark line.

Predictor	t_{sd}	$d(t_{sd})$	t^{m}_{sd}	$d(t^{m}_{sd})$
Intercept	3.54 [3.49 – 3.59]		3.14 [3.10 – 3.18]	
External temperature	-0.05 [-0.06 – -0.05]	-0.08	-0.04 [-0.04 – -0.04]	-0.08
Age: 50-59	-0.07 [-0.10 – -0.05]	-0.11	-0.08 [-0.10 – -0.06]	-0.13
Age: 60-69	-0.13 [-0.15 – -0.10]	-0.19	-0.13 [-0.15 – -0.11]	-0.23
Age: 70-79	-0.18 [-0.21 – -0.15]	-0.26	-0.16 [-0.19 – -0.13]	-0.28
Sex: Male	-0.06 [-0.07 – -0.04]	-0.08	-0.09 [-0.10 – -0.08]	-0.16
Ethnicity: Chinese	0.16 [0.05 – 0.28]	0.24	0.25 [0.15 – 0.35]	0.44
Ethnicity: Other ethnic group	0.04 [-0.04 – 0.12]	0.06	0.09 [0.02 – 0.16]	0.16
Tenure: Mortgage	-0.05 [-0.07 – -0.04]	-0.08	-0.03 [-0.04 – -0.01]	-0.05
Tenure: Rent Local Authority	-0.11 [-0.15 – -0.07]	-0.16	-0.09 [-0.12 – -0.05]	-0.16
Household size: two	-0.03 [-0.05 – -0.02]	-0.05	-0.06 [-0.08 – -0.05]	-0.11
Household size: three	-0.04 [-0.07 – -0.02]	-0.06	-0.06 [-0.08 – -0.04]	-0.10
Fuel: An Open solid fuel fire	0.08 [0.04 – 0.12]	0.12	0.05 [0.02 – 0.09]	0.09
Fuel: Gas Hob & Osf fire	0.10 [0.08 – 0.13]	0.15	0.06 [0.04 – 0.09]	0.11
Fuel: Gas Fire & Osf fire	0.11 [-0.02 – 0.24]	0.16	0.15 [0.04 – 0.27]	0.27
Fuel: Gas Hob &Gas Fire & Osf fire	0.11 [0.06 – 0.17]	0.17	0.11 [0.06 – 0.16]	0.19
BMI: underweight	0.12 [0.04 – 0.19]	0.17	0.11 [0.05 – 0.18]	0.20
BMI: overweight	-0.18 [-0.19 – -0.17]	-0.26	-0.14 [-0.15 – -0.13]	-0.24
BMI: obese	-0.36 [-0.41 – -0.30]	-0.52	-0.26 [-0.30 – -0.21]	-0.45
Activity: 20-40	0.13 [0.11 – 0.15]	0.20	0.04 [0.02 – 0.06]	0.07
Activity: 40-60	0.23 [0.22 – 0.25]	0.34	0.09 [0.07 – 0.11]	0.16
Activity: 60-80	0.32 [0.30 – 0.34]	0.46	0.11 [0.10 – 0.13]	0.20
Activity: 80-100	0.48 [0.46 – 0.50]	0.70	0.19 [0.17 – 0.20]	0.33
Health: Moderately happy	-0.10 [-0.12 – -0.07]	-0.14	-0.05 [-0.07 – -0.03]	-0.09
Health: Moderately unhappy	-0.15 [-0.19 – -0.12]	-0.22	-0.09 [-0.12 – -0.06]	-0.16
Health: Very unhappy	-0.16 [-0.21 – -0.11]	-0.23	-0.08 [-0.13 – -0.03]	-0.14
Health: Extremely unhappy	-0.15 [-0.23 – -0.07]	-0.22	-0.07 [-0.14 – -0.00]	-0.13
Heating: Oil central heating	0.09 [0.05 – 0.13]	0.13	0.07 [0.03 – 0.10]	0.11
Observations		37730		37730
R^2 / adjusted R^2		0.24		0.19

Table B.8: The effect size for the statistically significant differences for multiple linear regression models of thermal variety. Those marked in red have $p > 0.01$ and therefore are not considered significant. The effect size is calculated by dividing the estimate by the sample standard deviation, as discussed in chapter 6.5. $sd(t_{sd} = 0.69)$, $sd(t^m_{sd} = 0.57)$.

Continuous variables		Risk ratio (t_{sd})	Risk ratio (t^{m}_{sd})	Signif.
t_{sd}		0.95 [0.93 – 0.98]	NA	**
t^{m}_{sd}		NA	0.97 [0.94 – 1.00]	
Categorical variables	**subcategory**	**Risk ratio**	**Risk ratio**	
Age	40-49			
	50-59	1.48 [1.32 – 1.64]	1.48 [1.33 – 1.64]	***
	60-69	2.10 [1.88 – 2.34]	2.10 [1.89 – 2.34]	***
	70+	2.70 [2.41 – 3.03]	2.71 [2.41 – 3.04]	***
Sex	Female			
	Male	1.52 [1.47 – 1.58]	1.52 [1.47 – 1.58]	***
Ethnic background	White			
	Mixed	0.99 [0.74 – 1.32]	0.99 [0.74 – 1.32]	
	Asian or Asian British	1.16 [0.97 – 1.38]	1.16 [0.97 – 1.39]	
	Black or Black British	0.90 [0.71 – 1.15]	0.90 [0.71 – 1.14]	
	Chinese	0.86 [0.51 – 1.44]	0.86 [0.51 – 1.44]	
	Other ethnic group	1.12 [0.88 – 1.44]	1.12 [0.88 – 1.44]	
	Do not know	1.17 [0.41 – 3.39]	1.17 [0.41 – 3.39]	
	Prefer not to answer	0.95 [0.68 – 1.33]	0.95 [0.68 – 1.33]	
Household income	Less than 18,000			
	18,000 to 30,999	0.91 [0.86 – 0.96]	0.91 [0.86 – 0.96]	**
	31,000 to 51,999	0.80 [0.75 – 0.85]	0.80 [0.75 – 0.85]	***
	52,000 to 100,000	0.72 [0.67 – 0.77]	0.72 [0.67 – 0.77]	***
	Greater than 100,000	0.64 [0.57 – 0.71]	0.64 [0.58 – 0.71]	***
	Prefer not to say	0.84 [0.77 – 0.91]	0.84 [0.77 – 0.91]	**
	Do not know	0.92 [0.82 – 1.03]	0.92 [0.82 – 1.03]	
Tenure type	Own outright			
	None of above	0.91 [0.66 – 1.26]	0.92 [0.66 – 1.26]	
	Prefer not to answer	1.08 [0.78 – 1.49]	1.08 [0.78 – 1.49]	
	Mortgage	1.07 [1.03 – 1.12]	1.08 [1.03 – 1.13]	*
	Rent Local Authority	1.22 [1.10 – 1.35]	1.22 [1.11 – 1.35]	**
	Rent private	1.11 [0.98 – 1.26]	1.11 [0.98 – 1.26]	
	Shared	1.47 [1.07 – 2.01]	1.47 [1.07 – 2.02]	
	Rent free	1.04 [0.83 – 1.30]	1.04 [0.83 – 1.31]	
Accommodation type	House/bungalow			
	None of above	0.80 [0.45 – 1.41]	0.80 [0.45 – 1.41]	
	Prefer not to answer	1.18 [0.40 – 3.52]	1.19 [0.40 – 3.54]	
	Flat	0.97 [0.91 – 1.05]	0.98 [0.91 – 1.05]	
	Temporary	0.81 [0.41 – 1.62]	0.81 [0.41 – 1.62]	

Table B.9: Results of the binomial regression of C_{EWD} with the sociodemographic, housing and health factors described in the text. The total number of participants and the percentages for either $C_{EWD} = 0$ or $C_{EWD} = 1$ are given, along with the risk ratio and 95% confidence interval. The relative subcategory for each variable does not have an RR estimate. Two models are shown, one including t_{sd} and other t^{m}_{sd}. Significance levels: * $p < 0.01$, ** $p < 0.001$, *** $p < 1 \times 10^{-9}$. N=77,762. The p value for t^{m}_{sd} was 0.09.

Categorical variables	subcategory	Risk ratio (t_{sd})	Risk ratio (t^m_{sd})	Signif.
Employment status	In paid/self-employment			
	None of the above	1.1 [0.85 – 1.41]	1.1 [0.85 – 1.41]	
	Prefer not to answer	0.76 [0.42 – 1.36]	0.76 [0.42 – 1.36]	
	Retired	1.06 [1.01 – 1.11]	1.06 [1.01 – 1.11]	
	Looking after home and/or family	0.96 [0.86 – 1.08]	0.96 [0.86 – 1.08]	
	Unable to work	1.82 [1.66 – 1.99]	1.83 [1.67 – 2.00]	***
	Unemployed	0.85 [0.71 – 1.01]	0.85 [0.71 – 1.01]	
	Doing unpaid or voluntary work	1.06 [0.97 – 1.15]	1.06 [0.97 – 1.15]	
	Full or part-time student	1.04 [0.84 – 1.28]	1.04 [0.84 – 1.27]	
Fuel type	Gas hob or gas cooker			
	Gas fire	1.05 [0.99 – 1.12]	1.05 [0.99 – 1.12]	
	Open solid fuel (s.f) open fire	1.01 [0.91 – 1.13]	1.01 [0.90 – 1.12]	
	Gas hob & Gas Fire	1.05 [1.00 – 1.09]	1.05 [1.00 – 1.09]	
	Gas hob & s.f. open fire	0.92 [0.85 – 1.01]	0.92 [0.84 – 1.01]	
	Gas fire & s.f. open fire	1.05 [0.74 – 1.50]	1.04 [0.73 – 1.49]	
	Gas hob & Gas fire & s.f. open fire	0.98 [0.83 – 1.16]	0.98 [0.83 – 1.16]	
	None of the above	1.01 [0.96 – 1.06]	1.01 [0.96 – 1.06]	
	Prefer not to say	1.64 [0.95 – 2.85]	1.65 [0.95 – 2.87]	
	Do not know	1.52 [0.63 – 3.67]	1.53 [0.63 – 3.69]	
Body mass index	normal			
	underweight	1.00 [0.76 – 1.31]	1.00 [0.76 – 1.31]	
	overweight	1.15 [1.11 – 1.20]	1.16 [1.12 – 1.21]	***
	obese	1.49 [1.32 – 1.68]	1.51 [1.33 – 1.71]	***
Activity level during study week, by quintile, lowest to highest activity	1st quintile			
	2nd quintile	0.84 [0.80 – 0.89]	0.84 [0.80 – 0.88]	***
	3rd quintile	0.81 [0.77 – 0.86]	0.81 [0.77 – 0.85]	***
	4th quintile	0.77 [0.73 – 0.82]	0.76 [0.72 – 0.81]	***
	5th quintile	0.72 [0.68 – 0.77]	0.71 [0.67 – 0.76]	***
Household size	Single occupant			
	Two	1.13 [1.07 – 1.19]	1.13 [1.07 – 1.19]	**
	Three	1.20 [1.12 – 1.29]	1.20 [1.12 – 1.29]	**
	Four or more	1.18 [1.09 – 1.27]	1.18 [1.09 – 1.27]	**

Table B.10: A continuation of the results of the binomial regression shown in B.9 of C_{EWD} with the sociodemographic, housing and health factors described in the text. The total number of participants and the percentages for either $C_{EWD} = 0$ or $C_{EWD} = 1$ are given, along with the risk ratio and 95% confidence interval. The relative subcategory for each variable does not have an RR estimate. Two models are shown, one including t_{sd} and other t^m_{sd}. Significance levels: * $p < 0.01$, ** $p < 0.001$, *** $p < 1 \times 10^{-9}$. N=77,762.

Appendix C

Python scripts

C.1 Downsampling

The following script downsamples the 5-second temperature and activity data for each participant to a 1-minute period.

```
"""
Created on Wed Mar 21 11:13:21 2018

@author: Harry Kennard
"""
import pandas as pd
import os
import numpy as np
directory_in_str = r"R:\BioBank\results_apr14oct14"
directory = os.fsencode(directory_in_str)
suffix = os.fsencode("Epoch.csv")
suffixNonWear = os.fsencode("NonWearBouts.csv")
print("Running..")
for file in os.listdir(directory):
    filename = os.path.join(directory,file,file+suffix)
    filename = filename.decode("utf-8")
    filenameNonWear = os.path.join(directory,file,file+suffixNonWear)
    filenameNonWear = filenameNonWear.decode("utf-8")
    try:
        data = pd.read_csv(filename, engine='python', index_col=0)
        dataNonWear = pd.read_csv(filenameNonWear,engine='python')
        dataNonWear = dataNonWear.drop(['xStdMax', 'yStdMax', 'zStdMax'],axis
    =1)
        data.index = pd.to_datetime(data.index)
        data2=data.resample('1Min').mean()
        data2=data2.drop(['samples','dataErrors','clipsBeforeCalibr','
    clipsAfterCalibr','rawSamples'],axis=1)
        data2['wear']=np.ones(len(data2.index))
        data2['temp']=data2['temp'].round(2)
```

```
        len(data2.index)
        for index, row in dataNonWear.iterrows():
            start = row['start']
            end = row['end']
            data2['wear'].loc[start:end] = 0
        data2.to_csv(file.decode("utf-8") + ".csv", sep=',')
    except:
        print(filename + " failed")
print("done")
```

C.2 Metric processing

The following script produces the summary metrics which are used in this study.

```
"""
Created on Mon Jan 21 18:38:05 2019
@author: Kennard
"""
import pandas as pd
import os
import numpy as np
import csv
from scipy import stats as st
dir_list = []
dir_list.append(r"C:\BioBank\downsampled")

output_dir = r"C:\BioBank\multix"

def patient_id(cwa_filename):
    try:
        name = os.path.basename(cwa_filename)
        return int(name.split('_')[0])
    except:
        print("Not a valid id")
def find_csv_filenames(path_to_dir, suffix=".csv" ):
    filenames = os.listdir(path_to_dir)
    return [filename for filename in filenames if filename.endswith(suffix)]
def list_of_ids(input_dir):
    directory = os.fsencode(input_dir)
    suffix = os.fsencode(".csv")
    files = find_csv_filenames(directory,suffix)
    filelist = []
    for file in files:
        filename = os.path.join(directory,file)
        filename = filename.decode("utf-8")
        pat_id = patient_id(filename)
        filelist.append(str(pat_id))
```

```
    return filelist
for input_dir in dir_list:
    directory = os.fsencode(input_dir)
    output_med = os.path.join(os.fsencode(output_dir),os.fsencode(os.path.split(input_dir)[1]+"g.csv"))
    output_x = os.path.join(os.fsencode(output_dir),os.fsencode(os.path.split(input_dir)[1]+"f.csv"))
    output = os.path.join(os.fsencode(output_dir),os.fsencode(os.path.split(input_dir)[1]+"p.csv"))
    print("processing",directory)
    with open(output_med, 'w', newline='') as f:
      wr_med= csv.writer(f)
      wr_med.writerow(("eid","count","mean","std","min","x10","x25","x50","x75","x90","max"))
      with open(output_x, 'w', newline='') as g:
            wr_x= csv.writer(g)
            wr_x.writerow(("eid","count","mean","std","min","x10","x25","x50","x75","x90","max"))
            with open(output, 'w', newline='') as h:
                wr= csv.writer(h)
                file_list = find_csv_filenames(input_dir)
                wr.writerow(("eid","count","mean","std","min","x10","x25","x50","x75","x90","max"))
                for file in file_list:
                    file = os.fsencode(file)
                    filename = os.path.join(directory,file)
                    filename = filename.decode("utf-8")
                    pat_id = patient_id(filename)
                    try:
                        data = pd.read_csv(filename, engine='python', index_col=0)
                        data.index = pd.to_datetime(data.index)
                        grouped = data.groupby('wear')
                        g1 = grouped.get_group(1.0)
                        size = g1['temp'].describe()[0]
                        if size >9072:
                            drp = g1[1440:-1440] #remove first and last days
                            drp_med = drp[drp['enmoTrunc'] > 0.01]
                            drp_x = drp[drp['enmoTrunc'] < 0.01]
                            dat_med=np.round(drp_med['temp'].describe(percentiles = [0.1,0.25,0.5,0.75,0.9]), decimals = 3)
                            dat_x=np.round(drp_x['temp'].describe(percentiles = [0.1,0.25,0.5,0.75,0.9]), decimals = 3)
                            dat=np.round(drp['temp'].describe(percentiles = [0.1,0.25,0.5,0.75,0.9]), decimals = 3)
```

```
                    wr.writerow([pat_id,dat[0],dat[1],dat[2],dat[3],dat[4],dat[5],dat[6],dat[7],dat[8],dat[9]])
                    wr_x.writerow([pat_id,dat_x[0],dat_x[1],dat_x[2],dat_x[3],dat_x[4],dat_x[5],dat_x[6],dat_x[7],dat_x[8],dat_x[9]])
                    wr_med.writerow([pat_id,dat_med[0],dat_med[1],dat_med[2],dat_med[3],dat_med[4],dat_med[5],dat_med[6],dat_med[7],dat_med[8],dat_med[9]]) #write to file
                else:
                    print([pat_id, size])
            except (OSError, IOError,KeyError, ValueError, IndexError) as e:
                if e == 1.0:
                    print("error =",e,filename)
```

Appendix D

Heating savings calculation

The order of magnitude energy savings for a 1°C reduction in heating demand temperature are 10% (Research, 2012). This estimate likely holds for the EU in general because the UK represents the average heating degree days for EU28 countries (UK: 3179.4, EU28: 3217.6 in 2013) (Eurostat, 2019a) .

Restricting the analysis to natural gas provides a lower limit for the total domestic space heating emissions, since carbon intensities for natural gas combustion for heat are more readily available than other fuel types. EU final energy demand was 1,122 Mtoe in 2017 (Eurostat, 2019b). Total EU28 space heating demand from natural gas was 27.6% of residential energy demand in 2017, and residential demand itself was 27.2% of final energy demand (Eurostat, 2019c). Therefore, natural gas domestic heating demand was 84.2 Mtoe (9.80×10^{11} kWh) in 2017.

High efficiency boilers result in carbon emissions of between 210-230 gCO_2/kWh (CCC, 2013). This number will overestimate average European boiler efficiency, and therefore underestimate heating CO_2 emissions, and provide a lower limit for emissions overall. Using a value of 230 gCO_2/kWh results in total EU28 domestic heating emissions of at least 225 $MtCO_2e$ from natural gas alone (compare 4466 $MtCO_2e$ emissions for EU28 as a whole (EEA, 2019). Applying the estimate that 10% of these emissions would be curtailed by a 1°C reduction in domestic demand temperature yields 22.5 $MtCO_2e$/annum.

An upper limit is more challenging to derive, but total space heating proportion of final energy use provides an estimate. In 2017 this was 64.1%, which is 195.6 Mtoe (2.28×10^{12} kWh) (Eurostat, 2019c). Assuming an average carbon intensity of 300 gCO_2/kWh gives a balance between higher intensity oil and petroleum heating sources (which were 9% of residential end use in 2017) and lower intensity renewable heat sources (which were 15% of residential end use in 2017) yields an upper limit of 683 $MtCO_2e$, or 68.3 $MtCO_2e$/annum saving on a 1°C reduction of demand temperature. The average of the two limit estimates is approximately 45 $MtCO_2e$/annum.

Estimating the maximum proportion of the European population who might be able to tolerate a 1°C reduction in domestic temperatures is also challenging. 13.5% of the UK Biobank sample used in this study have conditions associated with excess winter deaths. While the UK Biobank sample is likely to be healthier than the general population (Fry et al., 2017), selecting the 25% of households which have fit and healthy inhabitants for a heating reduction campaign would be unlikely to greatly increased risk of harm. Therefore, approximately 10 $MtCO_2e$/annum emissions might be avoidable. Of course, this estimate does not take into account comfort considerations, willingness to participate and potential unforeseen consequences.

Appendix E

Pre–Analysis Plan

E.1 Introduction

This study is motivated by questions around the impacts of cold domestic conditions on health. In the UK, the primary evidence for a link between cold and poor health is the peak in mortality which occurs in winter, as recorded in excess winter mortality statistics. To date, there have been a limited number of large scale studies into the population level links between cold exposure and health. The study makes use of UK Biobank data, and in particular temperature data taken from an wristband activity monitor worn by over 100,000 Biobank participants for a week. Pilot studies showed that the temperature recorded by the wristband is a determined by a mixture of ambient environmental temperature and heat from the wrist a quantity which has been called 'experienced temperature' a concept adapted from Kuras et al. (2015). Following a large data processing exercise to down-sample the 100k data files to produce the experienced temperature time series, associations between demographics/health condition data and the experienced temperature will be tested. This document is intended to summarise the statistical processing which will be undertaken. Justification of other methodical choices will be outlined in other supporting documents and future publications.

E.1.1 Research Questions

1. Does experience temperature vary with demographic/household variables [e.g. sex, age, ethnicity, income]?

   Sub-question: Are there regional effects which are not explained by external temperature?

2. Are there associations between experienced temperature and the health conditions associated with excess winter deaths [ICD 10 I00-I99, J00J99 F01, F03, G30]?

3. Do combinations of demographics (i.e. low income) and health conditions associated with excess winter deaths [ICD 10 I00-I99, J00J99 F01, F03, G30] have associations

with low experienced temperature?

Sub-question: Are there regional effects which are not explained by external temperature?

E.1.2 Hypotheses

1. No significant differences in experienced will be measured as a function of demographics.

2. A weak association between low experienced temperature and health conditions associated with excess winter deaths is expected.

3. A stronger association between low experienced temperature and health conditions associated with excess winter deaths and low household income.

E.2 Data

The data used in this study are supplied by the UK Biobank, which is large ongoing longitudinal heath study based in the UK. As part of this study potential participants were invited to wear a wrist activity monitor. The monitor also records temperature, at a period of around 2 seconds. Participants were instructed to wear the wristband continuously for a week. Around 33% of invited participants produced data included in this study. The reasons for non-inclusion are detailed in figure E.2 below, and listed in greater detail in section E.3.2.

E.2.1 Excluded participants

- Participants with conditions related to cold hands such as anaemia, anorexia nervosa, Buerger's disease, carpal tunnel syndrome, diabetes, lupus, Raynaud's disease, scleroderma.

- Participants who carry out night shift work whose circadian rhythms are likely to be disrupted.

- Participants with conditions related to severe disruption of circadian rhythms, i.e. Alzheimer's disease and dementia.

E.2.2 Statistical significance

The alpha significance level for this study will be taken as 0.05. Power estimations in multilevel models are not straightforward, especially prior to the analysis stage. However, general comments regarding Cohen's d effect sizes may be made (Cohen, 1988). Given the standard deviation of the experienced temperature is 2.1°C (see figure 2), any between group differences of a similar order would represent a large effect size, according to Cohen's scale.

At the other end of the effect size scale, the temperature sensitivity of the wrist monitor is around 0.3°C group differences of this magnitude would correspond to an effect size of around 0.1, which is small on Cohen's scale.

It is expected that the total share of the variance that the variables in the model are able to explain will be low. This is due to a combination of incomplete variable specification, significant error in the variables that have been recorded, and many unknown-unknowns which contribute to the variability of social-scientific systems in general.

E.3 Empirical Analysis

The following section details the specific plan for the statistical analysis of the data, and the variables which will be included.

E.3.1 Experienced temperature

The experienced temperature is derived in the following manner (see figure 1 for the data processing outline). First, the temperature time series, at 5 second period, is divided into worn and unworn periods using the algorithm defined by Doherty et al. (2017). The data is then down-sampled to a period of a minute. Since this algorithm for wear/non-wear detection relies on activity detection, it is not completely reliable. The data is therefore first screened for a total 'wear time' of over 9072 minutes, or 90% of the total week. The first and last days are then removed, as they typically include transient periods when the monitor is moved around but not necessarily worn for example during unpacking from the postage envelope. Next, data with activity above the median is removed: these correspond to periods of activity such as exercise or travelling when a participant is unlikely to be in a sedentary domestic/work setting (it is not possible to accurately infer met levels from activity alone (Hees et al., 2011)). Finally, temperature data are ordered and the value corresponding to the lowest decile taken as the experienced temperature.

E.3.2 Variables

The UK Biobank has a very large number of participant variables available. The application for this study requested around 80. Of these, the 16 below were selected to best answer the research questions posed above. Limiting the number of variables reduces the probability of type I errors. Several variables are used in conjunction with each other (such as home location co-ordinates), as well as information not available in the UK Biobank (such as NASA external temperature data) to yield a total of 11 variables which enter the model.

E.3.2.1 Variables RQ1

1. 54. $\rightarrow$ Assessment centre (over 21 regions, given the index j)
2. 90100. $\rightarrow y_1$. Experienced temperature (continuous)

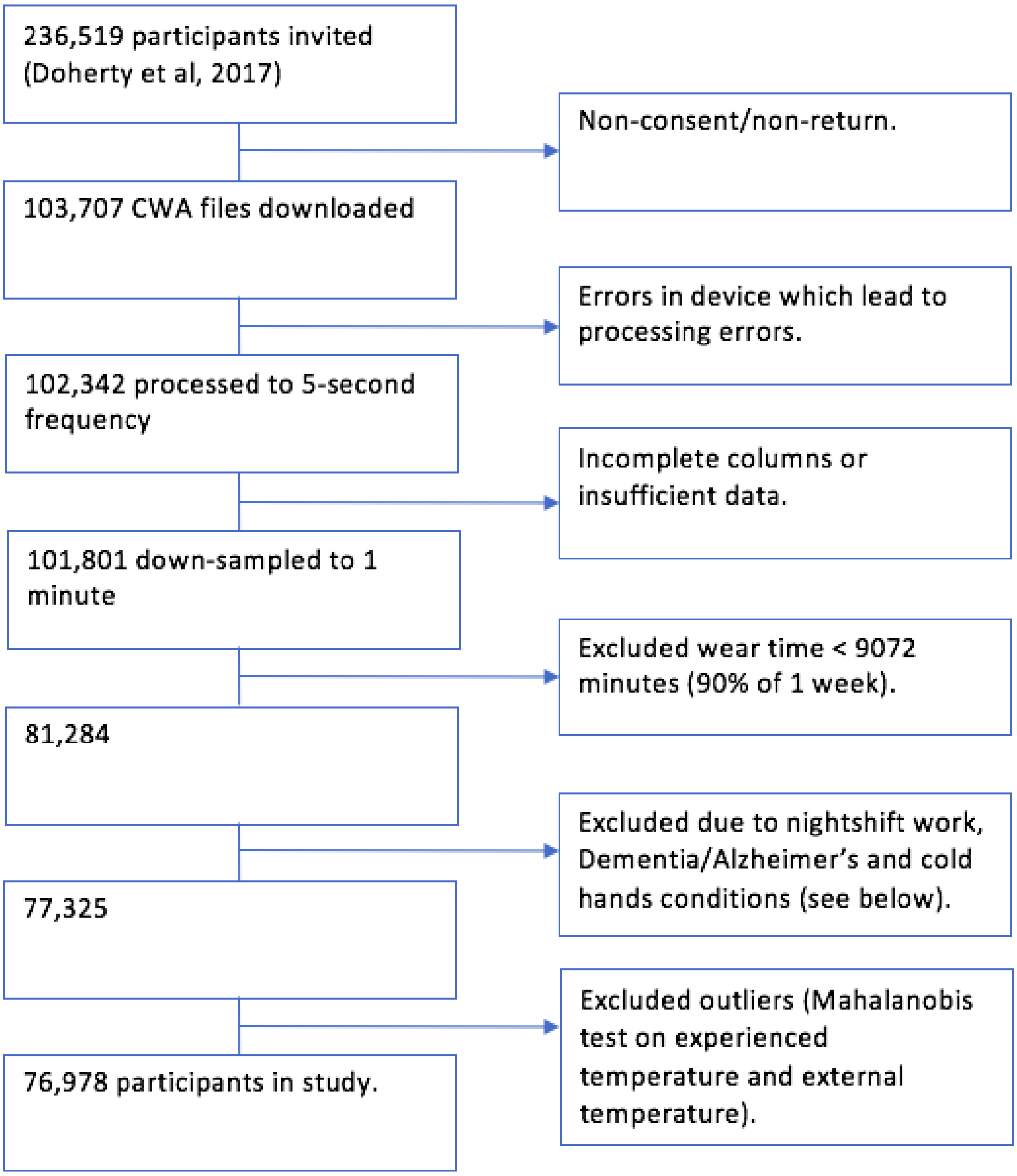


Figure E.1: The process by which variables were removed from the study

3. 20074, 20075 (home location) + NASA data $\rightarrow x_1$. Gridded external temperature (continuous)

4. 34./52. $\rightarrow x_2$. Age (continuous)

5. 31.$\rightarrow x_3$. Sex (binary)

6. 6139+6140 $\rightarrow x_4$: Heating type (categorical)

7. 670+680 $\rightarrow x_5$ Type of Housing + tenure (categorical)

8. 738—709. $\rightarrow x_6$. Household income normalised by number in household (category)

9. 20119. $\rightarrow x_7$ Current Employment status (categorical)

10. 4581. $\rightarrow x_8$ Financial situation satisfaction (categorical)

11. 21000. $\rightarrow x_9$. Ethnic background (categorical)

Co-variates (indicated by below): Age (x_2) and Employment Status (x_7), Household income (x_6) and Financial situation satisfaction (x_8)

E.3.2.2 Variables RQ2

1. 20002 + 4006. $\rightarrow z$ Excess winter death related illness (compound of ICD 10 J00J99 F01, F03, G30 and self-reporting)

2. 90100. $\rightarrow x$. Experienced temperature (continuous)

E.3.3 Variables RQ3

1. 54. $\rightarrow$ Assessment centre (over 21 regions, given the index j)

2. 20002+ 4006. $\rightarrow z_1$ Excess winter death related illness (combination of ICD 10 J00J99 F01, F03, G30 and self-reporting)

3. 90100. $\rightarrow x_1$:Experienced temperature

4. 20074, 20075 (home location) + NASA data $\rightarrow x_2$. Gridded External temperature (continuous)

5. 34./52. $\rightarrow x_3$. Age (continuous)

6. 31. $\rightarrow x_4$. Sex (binary)

7. 6139+6140 $\rightarrow x_5$: Heating type (categorical)

8. 670+680 $\rightarrow x_6$ Type of Housing + tenure (categorical)

9. 738—709. $\rightarrow x_7$. Household income normalised by number in household (category)

10. 20119. $\rightarrow x_8$ Current Employment status (categorical)

11. 4581. $\rightarrow x_9$ Financial situation satisfaction (categorical)

12. 21000. $\rightarrow x_{10}$. Ethnic background (categorical)

Co-variates (indicated by $\times$ below): Age (x_3) and Employment Status (x_8), Household income (x_7) and Financial situation satisfaction (x_9)

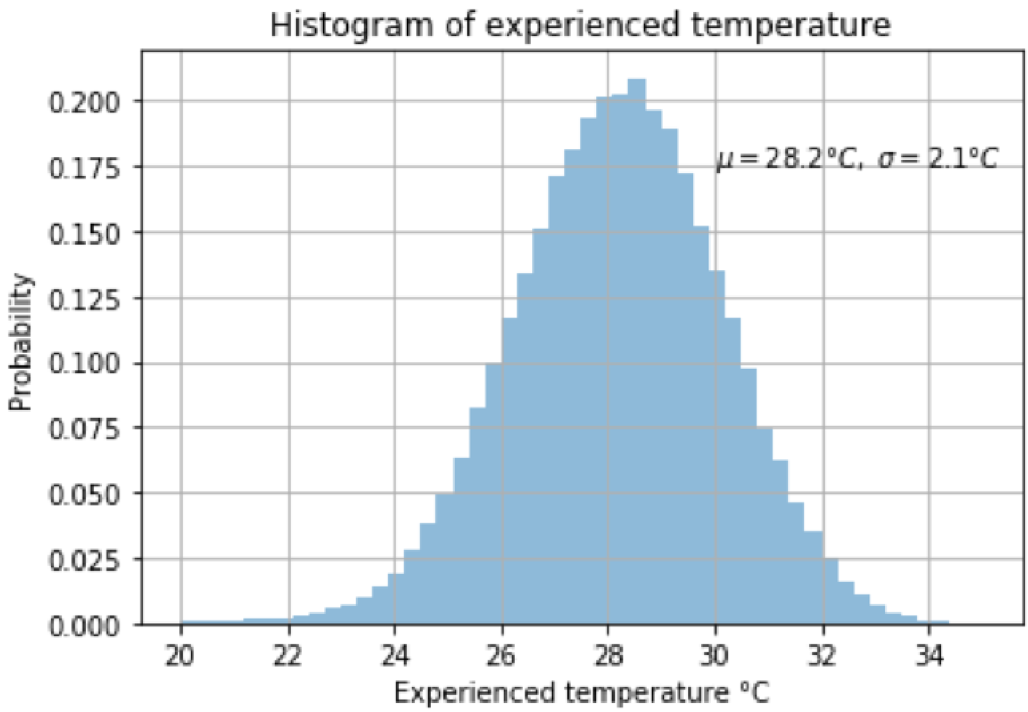


Figure E.2: A histogram of the processed experienced temperature. The data are approximately normally distributed.

E.3.4 Checks

For missing values, it is assumed the data are missing at random. Since the temperature and age variables occur on different numerical scales, the variables will be centred (by subtracting the mean) and standardised (by dividing by the standard deviation). For several variables, multiple instances are available, ranging between the initial assessment visit (2006-2010) and repeat visits between(2012-13). For all such variables the most recent available instance is used.

E.3.5 Multi-level model

Multilevel structure test (RQ1) Given the participants come from different geographical locations in the UK, which are naturally nested together in groups, a multilevel structure would seem appropriate. In order to understand if a multilevel structure is appropriate from a statistical standpoint, a test is needed to know if a null model differs significantly from a model with structure. The null single-level model is given by

$$y_{ij} = \beta_0 + e_{ij} \tag{E.1}$$

where y is the experienced temperature of the i^{th} participant in the j^{th} group, is the grand mean and the normally distributed residuals. The group-level model is given by

$$y_{ij} = \beta_0 + u_{0j} + e_{ij} \tag{E.2}$$

where u_j is the group random effect. A significant difference in these models indicates a multilevel structure is warranted. Such a difference is indicated by calculating the likelihood ratio test statistic (the 5% point of a chi-squared distribution on 1 d.f. is 3.84, a value greater than this suggests a significant difference between the models.) In the case that a multilevel model is appropriate, it will be given the following structure:

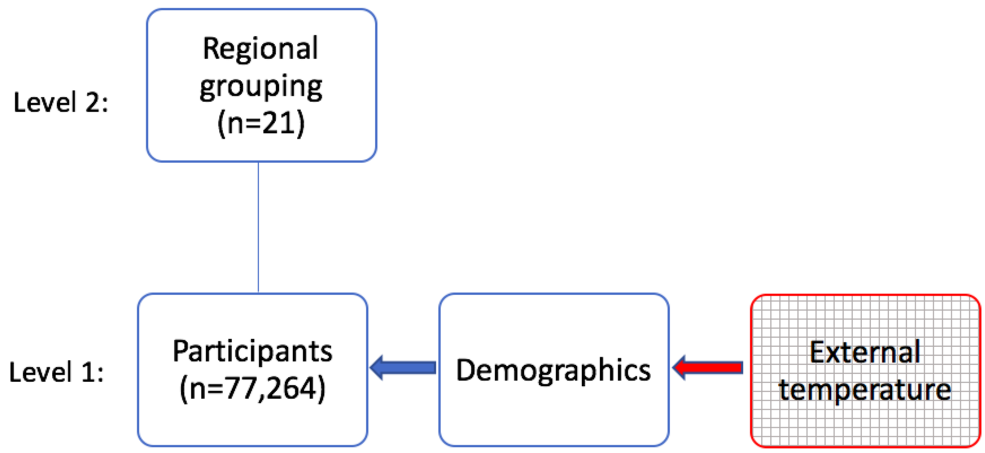


Figure E.3: A histogram of the processed experienced temperature. The data are approximately normally distributed.

This structure translates into the following regression equations for the i^{th} participant, j^{th} region and k^{th} variable. The regression will be built one by one, in the order given by the variable list above, adding a single variable at time to the model - the final form with all variables is as follows (N.B. the terms in each equation refer to the variables listed above):
RQ 1 equation

$$y_{ij} = \beta_0 + u_{0j} + \sum_{k=2}^{9} (\beta_k + u_{kj}) x_{k_{ij}} \times x_{1_{ij}} + e_{ij} \tag{E.3}$$

RQ 2 equation

$$z = \beta_0 + x + e_{ij} \tag{E.4}$$

RQ 3 equation

$$z_{ij} = \beta_0 + u_{0j} + (\beta_1 + u_{1j}) x_{1_{ij}} + \sum_{k=3}^{10} (\beta_k + u_{kj}) x_{k_{ij}} \times x_{2_{ij}} + e_{ij} \tag{E.5}$$

Practically, there will likely be a limit on the number of random slopes (u_{kj}) that can fitted with the data set, due to limited power. Since external temperature is likely to be the primary variable explaining experienced temperature, it takes precedent over the other variables in the model, and is inputted as a control for all variables. In the case of low power, the term u_{kj} inside the summation will be dropped.

E.3.6 Error Adjustments and other considerations

One of the advantages of multilevel modelling is that it correctly accounts for clustering in the data and provides robust error estimates. The assumptions that go into the method are addressed in Finch et al. (2014). The standard deviation of the primary variable, the experienced temperature, is 2.1°C the extent to this variation is explained by the model will be the subject of a great of scrutiny in the interpretation stage of this project. Careful consideration will be given to the exact nature of the measurand for each variable, and the extent to which uncertainty may mask explanatory variation. A summary of the total numbers of participants within each category of the variables is given in the appendix.

E.4 Software

The initial processing of the CWA files supplied by the UK Biobank was conducted using a modified version of the Python/Java code available through the OpenSource project. The majority of the subsequent data processing has been conducted using Python, the scripts will be made available after publication. For the statistical testing will be carried out using R, and specifically the lmer package as laid out by Finch et al. (2014) and Bristol University's Centre for Multilevel modelling course (Szmaragd and Leckie, 2011).

E.5 Deliverables

5 second activity/lux and temperature data-set, returned to UK BioBank. 3 research papers, potentially centred around three topics

- RQ1 and the definition of experienced temperature.
- the (potential) correlations between experienced temperature and health.
- the wider implications of the work to the fields of environmental health, energy and fuel poverty.

E.6 Budget

£2650 of research data programming support work has been paid for by the CEE budget.

This research was made possible by support from the EPSRC Centre for Doctoral Training in Energy Demand (LoLo), grant numbers EP/L01517X/1 and EP/H009612/1

Appendix F

Colophon

This document was created using LaTeX and BibTeX, composed with `Overleaf.com`. The following section gives key software tools and computing environments used throughout the study:

- Anaconda (for Python)
- Spyder (for Python)
- Github
- Atom
- Inkscape
- Epii-reviewer 4
- Rstudio with R version 3.5.1

R packages used:

- gridExtra_2.3
- sp_1.3-1
- mapproj_1.2.6
- maps_3.3.0
- forcats_0.4.0
- lubridate_1.7.4
- epiR_0.9-99
- survival_2.42-3
- viridis_0.5.1
- viridisLite_0.3.0
- yhat_2.0-0
- plyr_1.8.4
- sjlabelled_1.0.16
- sjmisc_2.7.7
- sjPlot_2.6.2
- car_3.0-2
- carData_3.0-2
- apaTables_2.0.5
- ggplot2_3.1.0
- nlme_3.1-137
- tools_3.5.1
- TMB_1.7.15
- backports_1.1.3
- rgdal_1.3-9
- R6_2.4.0
- lazyeval_0.2.1
- colorspace_1.4-0
- withr_2.1.2
- tidyselect_0.2.5
- mnormt_1.5-5
- emmeans_1.3.2
- curl_3.3
- compiler_3.5.1

- BiasedUrn_1.07
- sandwich_2.5-0
- labeling_0.3
- scales_1.0.0
- mvtnorm_1.0-8
- psych_1.8.12
- ggridges_0.5.1
- digest_0.6.18
- stringr_1.4.0
- foreign_0.8-70
- minqa_1.2.4
- rio_0.5.16
- stringdist_0.9.5.1
- pkgconfig_2.0.2
- lme4_1.1-20
- plotrix_3.7-4
- pwr_1.2-2
- rlang_0.3.1
- readxl_1.3.0
- rstudioapi_0.9.0
- generics_0.0.2
- zoo_1.8-4
- dplyr_0.8.0.1
- zip_2.0.0
- magrittr_1.5
- modeltools_0.2-22
- bayesplot_1.6.0
- Matrix_1.2-14
- Rcpp_1.0.0
- munsell_0.5.0
- abind_1.4-5
- prediction_0.3.6.2
- stringi_1.3.1
- multcomp_1.4-8
- yaml_2.2.0
- snakecase_0.9.2
- MASS_7.3-51.1
- grid_3.5.1
- parallel_3.5.1
- crayon_1.3.4
- lattice_0.20-35
- ggeffects_0.8.0
- haven_2.1.0
- splines_3.5.1
- sjstats_0.17.3
- hms_0.4.2
- knitr_1.21
- pillar_1.3.1
- boot_1.3-20
- estimability_1.3
- yacca_1.1.1
- codetools_0.2-15
- stats4_3.5.1
- glue_1.3.0
- data.table_1.12.0
- modelr_0.1.4
- nloptr_1.2.1
- miscTools_0.6-22
- cellranger_1.1.0
- gtable_0.2.0
- purrr_0.3.0
- tidyr_0.8.2
- assertthat_0.2.0
- xfun_0.5
- openxlsx_4.1.0
- coin_1.2-2
- xtable_1.8-3
- broom_0.5.1
- coda_0.19-2
- tibble_2.0.1
- glmmTMB_0.2.3
- TH.data_1.0-10

Bibliography

Parul Agarwal, M Afshar Alam, and Ranjit Biswas. Issues, challenges and tools of clustering algorithms. *IJCSI International Journal of Computer Science Issues*, 8(3):6, 2011.

Herman Aguinis, Ryan K. Gottfredson, and Harry Joo. Best-Practice Recommendations for Defining, Identifying, and Handling Outliers. *Organizational Research Methods*, 16(2): 270–301, April 2013. ISSN 1094-4281. doi: 10.1177/1094428112470848.

A. Ambrose and R. Marchand. The contemporary landscape of fuel poverty research. *Indoor and Built Environment*, 26(7):875–878, 2017. doi: 10.1177/1420326X17724914.

R. Areas, L. Duarte, and L. Menna-Barreto. Comparative analysis of rhythmic parameters of the body temperature in humans measured with thermistors and digital thermometers. *Biological Rhythm Research*, 37(5):419–424, 2006.

Axivity. AX3 datasheet, 2015. URL `https://axivity.com/files/resources/AX3_Data_Sheet.pdf`. Accessed = 2017-05-23.

Thom Baguley. Standardized or simple effect size: What should be reported? *British Journal of Psychology*, 100(3):603–617, August 2009. ISSN 00071269. doi: 10.1348/000712608X377117.

Mihaela Ioana Baritz, Diana Laura Cotoros, and Cristina Singer. Thermographic analysis of hand structure when subjected to controlled effort. In *2013 E-Health and Bioengineering Conference (EHB)*, pages 1–4, Iasi, 2013. IEEE. ISBN 978-1-4799-2373-1 978-1-4799-2372-4.

Amelia E. Barr. *Maids, Wives, and Bachelors.* Dodd Mead & co., New York, 1898.

BEIS. Energy consumption in the UK 2018, 2018a. URL `https://www.gov.uk/government/statistics/energy-consumption-in-the-uk`. Accessed = 2018-07-26.

BEIS. Energy trends march 2018, 2018b. URL `https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/712458/Energy_Trends_March_2018.pdf`. Accessed = 2019-07-24.

BEIS. 2017 UK greenhouse gas emissions, provisional figures, 2018c. URL https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/695930/2017_Provisional_Emissions_statistics_2.pdf. Accessed = 2019-07-24.

BEIS. Energy trends, March 2019. URL https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/791293/Energy_Trends_March_2019.pdf. Accessed = 2019-07-24.

Lars Berglund. Regression dilution bias: Tools for correction methods and sample size calculation. *Upsala Journal of Medical Sciences*, 117(3):279–283, August 2012. ISSN 0300-9734. doi: 10.3109/03009734.2012.668143.

Krishnan Bhaskaran, Shakoor Hajat, Andy Haines, Emily Herrett, Paul Wilkinson, and Liam Smeeth. Short term effects of temperature on risk of myocardial infarction in England and Wales: time series regression analysis of the Myocardial Ischaemia National Audit Project (MINAP) registry. *BMJ*, 341:c3823, August 2010. ISSN 0959-8138, 1468-5833. doi: 10.1136/bmj.c3823.

Ruth Bonita, Robert Beaglehole, and Tord Kjellström. *Basic epidemiology*. World Health Organization, Geneva, 2. ed edition, 2006. ISBN 978-92-4-154707-9. OCLC: 255900006.

Arthur Bowley. *Elementary Manual of Statistics*. P.S. King & Son, London, 1920.

M. Bracci, V. Ciarapica, A. Copertaro, M. Barbaresi, N. Manzella, M. Tomasetti, S. Gaetani, F. Monaco, M. Amati, M. Valentino, V. Rapisarda, and L. Santarelli. Peripheral skin temperature and circadian biological clock in shift nurses after a day off. *International Journal of Molecular Sciences*, 17(5), 2016.

Mario Brito and Tânia Sousa. World primary energy production & consumption 1900-2010: What can be learned from past trends? In *Proceedings of 12th International Conference on Energy for a Clean Environment*, Lisbon, July 2015. ICEER.

L. Brown. National radiation survey in the UK: Indoor occupancy factors. *Radiation Protection Dosimetry*, 5(4):203208, 1983. ISSN 0144-8420. doi: 10.1093/oxfordjournals.rpd.a082694.

C. Camargo, J. Ordorica, E. J. De la Vega, J. E. Olguín, O. R. López, and J. A. López. Analysis of temperature on the surface of the wrist due to repetitive movements using sensory thermography. *Work*, 41 Suppl 1:2569–75, 2012.

M Campos, A Gomariz, M Balsa, M. A. Rol, J. A. Madrid, and F. J. García. Age classification through the evaluation of circadian rhythms of wrist temperature. In

CAEPIA 2016: Advances in Artificial Intelligence pp 99-109, volume 9868 LNAI, pages 99–109, 2016.

Stephanie Carreiro, David Smelson, Megan Ranney, Keith J. Horvath, R. W. Picard, Edwin D. Boudreaux, Rashelle Hayes, and Edward W. Boyer. Real-Time Mobile Detection of Drug Use with Wearable Biosensors: A Pilot Study. *Journal of Medical Toxicology*, 11 (1):73–79, 2015.

Stephanie Carreiro, Kelley Wittbold, Premananda Indic, Hua Fang, Jianying Zhang, and Edward W. Boyer. Wearable Biosensors to Detect Physiologic Change During Opioid Use. *Journal of Medical Toxicology*, 12(3):255–262, 2016.

P. Carrer, M. Maroni, D. Alcini, and D. Cavallo. Allergens in indoor air: Environmental assessment and health effects. *Science of the Total Environment*, 270(13):3342, 2001. doi: 10.1016/S0048-9697(00)00791-9.

Jimmy Carter. Energy and the national goals - a crisis of confidence, 07 1979. URL `https://www.americanrhetoric.com/speeches/jimmycartercrisisofconfidence.html`.

CCC. Reducing the uks carbon footprint. Technical report, Committee on Climate Change, London, 2013. URL `https://www.theccc.org.uk/wp-content/uploads/2013/04/Reducing-carbon-footprint-report.pdf`.

Jonathan D. Chambers and Tadj Oreszczyn. Deconstruct: A scalable method of as-built heat power loss coefficient inference for UK dwellings using smart meter data. *Energy and Buildings*, 183:443–453, January 2019. ISSN 0378-7788. doi: 10.1016/j.enbuild.2018.11.016.

Jian Cheng, Zhiwei Xu, Rui Zhu, Xu Wang, Liu Jin, Jian Song, and Hong Su. Impact of diurnal temperature range on human health: a systematic review. *International Journal of Biometeorology*, 58(9):2011–2024, November 2014. ISSN 1432-1254. doi: 10.1007/s00484-014-0797-5.

R. P. Chilcott and R. Farrar. Biophysical measurements of human forearm skin in vivo: Effects of site, gender, chirality and time. *Skin Research and Technology*, 6(2):64–69, 2000.

J. H. Choi and V. Loftness. Investigation of human body skin temperatures as a bio-signal to indicate overall thermal sensations. *Building and Environment*, 58:258–269, 2012.

J. H. Choi and D. Yeom. Study of data-driven thermal sensation prediction model as a function of local body skin temperatures in a built environment. *Building and Environment*, 121:130–147, 2017.

Jacob Cohen. *Statistical power analysis for the behavioral sciences.* L. Erlbaum Associates, Hillsdale, N.J, 2nd ed edition, 1988. ISBN 978-0-8058-0283-2.

K. J. Collins, J. C. Easton, Hazel Belfield-Smith, A. N. Exton-Smith, and R. A. Pluck. Effects of age on body temperature and blood pressure in cold environments. *Clinical Science*, 69(4):465–470, October 1985. ISSN 0143-5221, 1470-8736. doi: 10.1042/cs0690465. URL `http://www.clinsci.org/content/69/4/465`.

M. D. Corbalán-Tutau, J. A. Madrid, J.M. Ordovás, C. E. Smith, F. Nicolas, and M. Garaulet. Differences in daily rhythms of wrist temperature between obese and normal-weight women: associations with metabolic syndrome features. *Chronobiology international*, 28(5):425–33, May 2011.

M. D. Corbalán-Tutau, P. Gómez-Abellán, J. A. Madrid, M Canteras, J. M. Ordovás, and M Garaulet. Toward a chronobiological characterization of obesity and metabolic syndrome in clinical practice. *Clinical nutrition (Edinburgh, Scotland)*, 34(3):477–83, June 2015.

Germaine Cornelissen. Cosinor-based rhythmometry. *Theoretical Biology & Medical Modelling*, 11:16, April 2014.

B. A. Cunha and R. P. Tu. Fever in the neurosurgical patient. *Heart & lung : the journal of critical care*, 17(6 Pt 1):608–611, November 1988.

H. A. M. Daanen. Finger cold-induced vasodilation: a review. *European Journal of Applied Physiology*, 89(5):411–426, June 2003.

Hein A. M. Daanen and Wouter D. Van Marken Lichtenbelt. Human whole body cold adaptation. *Temperature*, 3(1):104–118, January 2016. ISSN 2332-8940. doi: 10.1080/23328940.2015.1135688.

Michael Daly. Association of ambient indoor temperature with body mass index in england. *Obesity*, 22(3):626629, 2014. ISSN 1930-739X. doi: 10.1002/oby.20546.

Z Daraktchieva. *Review and Update of Occupancy Factors for UK homes.* Public Health England, London, UK, 2018. URL `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/763306/review_and_update_of_occupancy_factors_for_UK_homes.pdf`.

Huw Talfryn Oakley Davies, Iain Kinloch Crombie, and Manouche Tavakoli. When can odds ratios mislead? *BMJ : British Medical Journal*, 316(7136):989–991, March 1998. ISSN 0959-8138.

Steven J. Davis and Ken Caldeira. Consumption-based accounting of CO2 emissions. *Proceedings of the National Academy of Sciences*, 107(12):5687–5692, March 2010. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0906974107. URL http://www.pnas.org/content/107/12/5687.

DECC. Fuel poverty detailed tables: 2013, 2015. URL https://www.gov.uk/government/statistics/fuel-poverty-detailed-tables-2013.

DEFRA. Rural population 2014/15, 2019. URL https://www.gov.uk/government/publications / rural-population-and-migration / rural-population-201415. Accessed = 2019-07-24.

Department of Health. Health and Winter Warmth. Technical report, Regional Public Health Group in the South East, 2007. URL http://southwest-ra.gov.uk/media/SWHB/newsletters/WinterCold_bulletin.pdf.

Olivier Deschenes. Temperature, human health, and adaptation: A review of the empirical literature. *Energy Economics*, 46:606–619, November 2014. ISSN 0140-9883. doi: 10.1016/j.eneco.2013.10.013.

Ana V. Diez-Roux. Multilevel Analysis in Public Health Research. *Annual Review of Public Health*, 21(1):171–192, May 2000. ISSN 0163-7525, 1545-2093. doi: 10.1146/annurev.publhealth.21.1.171.

Aiden Doherty, Dan Jackson, Nils Hammerla, Thomas Pltz, Patrick Olivier, Malcolm H. Granat, Tom White, Vincent T. van Hees, Michael I. Trenell, Christoper G. Owen, Stephen J. Preece, Rob Gillions, Simon Sheard, Tim Peakman, Soren Brage, and Nicholas J. Wareham. Large Scale Population Assessment of Physical Activity Using Wrist Worn Accelerometers: The UK Biobank Study. *PLOS ONE*, 12(2):e0169649, February 2017. ISSN 1932-6203. doi: 10.1371/journal.pone.0169649.

G. C. Donaldson and W. R. Keatinge. Excess winter mortality: influenza or cold stress? Observational study. *BMJ*, 324(7329):89–90, January 2002. ISSN 0959-8138, 1468-5833. doi: 10.1136/bmj.324.7329.89.

J. G. Eby and G. A. Colditz. Obesity/Overweight: Prevention and Weight Management. In Harald Kristian (Kris) Heggenhougen, editor, *International Encyclopedia of Public Health*, pages 602–609. Academic Press, Oxford, January 2008. ISBN 978-0-12-373960-5. doi: 10.1016/B978-012373960-5.00424-X.

EEA. Total greenhouse gas emissions and projections. Technical report, European Environment Agency, Copenhagen, 2019. URL https://www.eea.europa.eu/data-and-maps/indicators/greenhouse-gas-emission-trends-6/assessment-2.

EIA. *US Energy Information, Annual Energy Review 2011.* US Energy Information Administration, Washington DC, 2012. URL https://www.eia.gov/totalenergy/data/annual/pdf/aer.pdf. Accessed = 2019-05-30.

Eurostat. Heating degree days - monthly data. Technical report, Eurostat, Brussels, 2019a. URL http://ec.europa.eu/eurostat/documents/38154/4956229/HDD-2013.xlsx/51ddb6cb-2d64-4b07-8e5b-5f07908fdd63.

Eurostat. Energy consumption in 2017. Technical report, Eurostat, Brussels, 2019b. URL https://ec.europa.eu/eurostat/documents/2995521/9549144/8-07022019-AP-EN.pdf/4a5fe0b1-c20f-46f0-8184-e82b694ad492.

Eurostat. Energy consumption in households 2017. Technical report, Eurostat, Brussels, 2019c. URL https://ec.europa.eu/eurostat/statistics-explained/index.php/Energy_consumption_in_households#Energy_products_used_in_the_residential_sector.

L. R. C. Ferreira, M. A. L. Miguel, M. M. Figueiredo De Martino, and L Menna-Barreto. Circadian rhythm of wrist temperature and night shift-work. *Biological Rhythm Research*, 44(5):737–744, 2013.

Davide Filingeri. Neurophysiology of Skin Thermal Sensations. In Ronald Terjung, editor, *Comprehensive Physiology*, pages 1429–1491. John Wiley & Sons, Inc., Hoboken, NJ, June 2016. ISBN 978-0-470-65071-4. doi: 10.1002/cphy.c150040.

W. Holmes Finch, Jocelyn E. Bolin, and Ken Kelley. *Multilevel Modeling Using R.* CRC Press, Boca Raton, FL, 1 edition edition, June 2014. ISBN 978-1-4665-1585-7.

J Fox and G. Monette. Generalized collinearity diagnostics. *JASA*, 87:178183, 1992.

L. J. French, M. J. Camilleri, N. P. Isaacs, and A. R. Pollard. Temperatures and heating energy in new zealand houses from a nationally representative studyheep. *Energy and Buildings*, 39(7):770782, Jul 2007. ISSN 0378-7788. doi: 10.1016/j.enbuild.2007.02.004.

Anna Fry, Thomas J. Littlejohns, Cathie Sudlow, Nicola Doherty, Ligia Adamska, Tim Sprosen, Rory Collins, and Naomi E. Allen. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *American Journal of Epidemiology*, 186(9):1026–1034, November 2017. ISSN 0002-9262. doi: 10.1093/aje/kwx246.

A. Pharo Gagge and Yasunobu Nishi. Heat Exchange Between Human Skin Surface and Thermal Environment. In Ronald Terjung, editor, *Comprehensive Physiology*. John Wiley

& Sons, Inc., Hoboken, NJ, January 2011. ISBN 978-0-470-65071-4. doi: 10.1002/cphy.cp090105.

Elizabeth Gaskell. *The Life of Charlotte Brontë*. Smith, Elder & Co., London, reissue edition edition, 1857.

A Gasparrini, B Armstrong, and M G Kenward. Distributed lag non-linear models. *Statistics in Medicine*, 29(21):22242234, Sep 2010. ISSN 0277-6715. doi: 10.1002/sim.3940.

A. Gasparrini, Y. Guo, F. Sera, A.M. Vicedo-Cabrera, V. Huber, S. Tong, M. de Sousa Zanotti Stagliorio Coelho, P.H. Nascimento Saldiva, E. Lavigne, P. Matus Correa, and et al. Projections of temperature-related excess mortality under climate change scenarios. *The Lancet Planetary Health*, 1:e360e367, 2017. doi: 10.1016/S2542-5196(17)30156-0.

Antonio Gasparrini and Michela Leone. Attributable risk from distributed lag models. *BMC Medical Research Methodology*, 14(1):55, Apr 2014. ISSN 1471-2288. doi: 10.1186/1471-2288-14-55.

Antonio Gasparrini, Yuming Guo, Masahiro Hashizume, Eric Lavigne, Antonella Zanobetti, Joel Schwartz, Aurelio Tobias, Shilu Tong, Joacim Rocklv, Bertil Forsberg, Michela Leone, Manuela De Sario, Michelle L Bell, Yue-Liang Leon Guo, Chang-fu Wu, Haidong Kan, Seung-Muk Yi, Micheline de Sousa Zanotti Stagliorio Coelho, Paulo Hilario Nascimento Saldiva, Yasushi Honda, Ho Kim, and Ben Armstrong. Mortality risk attributable to high and low ambient temperature: a multicountry observational study. *The Lancet*, 386 (9991):369–375, July 2015. ISSN 0140-6736. doi: 10.1016/S0140-6736(14)62114-0.

S. Gauthier. Investigating the probability of behavioural responses to cold thermal discomfort. *Energy and Buildings*, 124:7078, 2016. doi: 10.1016/j.enbuild.2016.04.036.

Global Modeling and Assimilation Office GMAO. inst1_2d_asm_Nx: MERRA-2 3D IAU State, Meteorology Instantaneous 3-hourly (p-coord, 0.625x0.5L42), version 5.12.4. Technical report, NASA, 2015.

Global Modeling and Assimilation Office GMAO. MERRA-2: File Specification Office Note No. 9 (Version 1.1). Technical report, NASA, 2016. URL `https://gmao.gsfc.nasa.gov/pubs/docs/Bosilovich785.pdf`.

Geoff Green and Jan Gilbertson. Warm front, better health. Technical report, Centre for Regional, Social and Economic Research, Sheffield, May 2008.

Steven A. Greenberg. How citation distortions create unfounded authority: analysis of a citation network. *BMJ*, 339:b2680, July 2009. ISSN 0959-8138, 1468-5833. doi: 10.1136/bmj.b2680.

Jae H. Kim. How to choose the level of significance: A pedagogical note. *SSRN Electronic Journal*, Jan 2015. doi: 10.2139/ssrn.2652773.

S. Hajat and A. Haines. Associations of cold temperatures with GP consultations for respiratory and cardiovascular disease amongst the elderly in London. *International Journal of Epidemiology*, 31(4):825–830, August 2002. ISSN 0300-5771. doi: 10.1093/ije/31.4.825. URL `https://academic.oup.com/ije/article/31/4/825/630272`.

Shakoor Hajat. Health effects of milder winters: a review of evidence from the United Kingdom. *Environmental Health*, 16(1):109, December 2017. ISSN 1476-069X. doi: 10.1186/s12940-017-0323-4.

Shakoor Hajat and Antonio Gasparrini. The Excess Winter Deaths Measure: Why Its Use Is Misleading for Public Health Understanding of Cold-related Health Impacts. *Epidemiology*, 27(4):486, July 2016. ISSN 1044-3983. doi: 10.1097/EDE.0000000000000479.

I. G. Hamilton, A. O' Sullivan, G. Huebner, T. Oreszczyn, D. Shipworth, A. Summerfield, and M. Davies. Old and cold? Findings on the determinants of indoor temperatures in English dwellings during cold conditions. *Energy and Buildings*, 141:142–157, 2017.

J. Hansen, M. Sato, P. Kharecha, and K. von Schuckmann. Earth's energy imbalance and implications. *Atmospheric Chemistry and Physics*, 11(24):13421–13449, December 2011. ISSN 1680-7316.

Brianna D. Harfmann, Elizabeth A. Schroder, Jonathan H. England, Natalie J. Senn, Philip M. Westgate, Karyn A. Esser, and Philip A. Kern. Temperature as a Circadian Marker in Older Human Subjects: Relationship to Metabolic Syndrome and Diabetes. *Journal of the Endocrine Society*, 1(7):843–851, 2017.

J. D. Healy. Excess winter mortality in Europe: a cross country analysis identifying key risk factors. *Journal of Epidemiology & Community Health*, 57(10):784–789, October 2003. ISSN 0143-005X, 1470-2738. doi: 10.1136/jech.57.10.784.

Vincent T. van Hees, Frida Renstrm, Antony Wright, Anna Gradmark, Michael Catt, Kong Y. Chen, Marie Lf, Les Bluck, Jeremy Pomeroy, Nicholas J. Wareham, Ulf Ekelund, Sren Brage, and Paul W. Franks. Estimation of Daily Energy Expenditure in Pregnant and Non-Pregnant Women Using a Wrist-Worn Tri-Axial Accelerometer. *PLOS ONE*, 6(7):e22922, July 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0022922.

H. Hensel and K. Schafer. Thermoreception and Temperature Regulation in Man. In E. Francis J. Ring and Barbara Phillips, editors, *Recent Advances in Medical Thermology*, pages 51–64. Springer New York, Boston, MA, 1984. ISBN 978-1-4684-7697-2. doi: 10.1007/978-1-4684-7697-2_8.

John Hills. Getting the measure of fuel poverty: Final Report of the Fuel Poverty Review. Technical report, Centre for Analysis of Social Exclusion, The London School of Economics and Political Science, 2012. URL `http://eprints.lse.ac.uk/43153`.

T. H. Howell. The skin temperature gradient in the upper limbs of old women. *Experimental gerontology*, 18(4):255–61, 1983.

HSE. *Workplace health, safety and welfare: A short guide for managers*. HSE Books, London, Dec 2007. ISBN 9780717662777. URL `http://www.hse.gov.uk/pubns/indg244.pdf`.

G. M. Huebner, M. Fell, M. Nicolson, D. Shipworth, S. Elam, C. Hanmer, H. Kennard, and C. Johnson. Are we heading towards a replicability crisis in energy efficiency research? a toolkit for improving the quality, transparency and replicability of energy efficiency impact evaluations. In *ECEEE Summer Study*, Belambra, 2017. ECEEE.

Gesche M. Huebner, Megan McMichael, David Shipworth, Michelle Shipworth, Mathieu Durand-Daubin, and Alex Summerfield. The reality of English living rooms A comparison of internal temperatures against common model assumptions. *Energy and Buildings*, 66: 688–696, November 2013. ISSN 0378-7788. doi: 10.1016/j.enbuild.2013.07.025.

Gesche M. Huebner, Megan McMichael, David Shipworth, Michelle Shipworth, Mathieu Durand-Daubin, and Alex J. Summerfield. The shape of warmth: temperature profiles in living rooms. *Building Research & Information*, 43(2):185–196, March 2015. ISSN 0961-3218. doi: 10.1080/09613218.2014.922339.

Gesche M. Huebner, Ian Hamilton, Zaid Chalabi, David Shipworth, and Tadj Oreszczyn. Comparison of indoor temperatures of homes with recommended temperatures and effects of disability and age: an observational, cross-sectional study. *BMJ Open*, 8(5):e021085, May 2018. ISSN 2044-6055, 2044-6055. doi: 10.1136/bmjopen-2017-021085.

Victor Hugo. *Les Misérables*. A. Lacroix, Verboeckhoven & Cie., Belgium, 1862.

Anders Huitfeldt. Is caviar a risk factor for being a millionaire? *BMJ*, 355:i6536, Dec 2016. ISSN 1756-1833. doi: 10.1136/bmj.i6536.

Jennifer A. Hutcheon, Arnaud Chiolero, and James A. Hanley. Random measurement error and regression dilution bias. *BMJ*, 340:c2289, June 2010. ISSN 0959-8138, 1468-5833. doi: 10.1136/bmj.c2289. URL `https://www.bmj.com/content/340/bmj.c2289`.

IEA. Global Energy & CO2 Status Report 2017. Technical report, IEA, 2018. URL http://www.iea.org/publications/freepublications/publication/GECO2017.pdf.

William G. Jacoby. Loess:: a nonparametric, graphical tool for depicting relationships between variables. *Electoral Studies*, 19(4):577–613, December 2000. ISSN 0261-3794. doi: 10.1016/S0261-3794(99)00028-1.

C. M. Jacquot, L Schellen, B. R. Kingma, M. A. van Baak, and W. D. van Marken Lichtenbelt. Influence of thermophysiology on thermal behavior: the essentials of categorization. *Physiology & behavior*, 128:180–7, April 2014.

T. W. Jang, H. Kim, S. H. Kang, S. H. Choo, I. S. Lee, and K. H. Choi. Circadian Rhythm of Wrist Temperature among Shift Workers in South Korea: A Prospective Observational Study. *International journal of environmental research and public health*, 14(10), September 2017.

R. Jevons, C. Carmichael, A. Crossley, and A. Bone. Minimum indoor temperature threshold recommendations for English homes in winter A systematic review. *Public Health*, 136: 4–12, July 2016.

F. Johnson, A. Mavrogianni, M. Ucci, A. Vidal-Puig, and J. Wardle. Could increased time spent in a thermal comfort zone contribute to population increases in obesity? *Obesity Reviews*, 12(7):543551, 2011. ISSN 1467-789X. doi: 10.1111/j.1467-789X.2010.00851.x.

British Medical Journal. Mortality in Cold Winters. *Br Med J*, 1(5327):347–348, February 1963. ISSN 0007-1447, 1468-5833. doi: 10.1136/bmj.1.5327.347. URL https://www.bmj.com/content/1/5327/347.

S K Firth and Andrew Wright. Investigating the thermal characteristics of english dwellings: Summer temperatures. *Proceedings of Conference: Air Conditioning and the Low Carbon Cooling Challenge - Windsor 2008 Conference*, Jan 2008.

Tom Kane. *Indoor temperatures in UK dwellings: investigating heating practices using field survey data*. PhD thesis, Loughborough University, 2013. URL https://dspace.lboro.ac.uk/2134/12563.

Alboukadel Kassambara. *Machine Learning Essentials: Practical Guide in R*. CreateSpace Independent Publishing Platform, Frankreich, 1 edition edition, March 2018. ISBN 978-1-986406-85-7.

W. R. Keatinge and G. C. Donaldson. Cold exposure and winter mortality in Europe. *The Lancet*, 350(9077):591–592, August 1997. ISSN 0140-6736. doi: 10.1016/S0140-6736(05)63181-9.

William R. Keatinge, Malcolm B. McIlroy, and Alan Goldfien. Cardiovascular responses to ice-cold showers. *Journal of Applied Physiology*, 19(6):1145–1150, November 1964. ISSN 8750-7587, 1522-1601.

Scott Kelly, Michelle Shipworth, David Shipworth, Michael Gentry, Andrew Wright, Michael Pollitt, Doug Crawford-Brown, and Kevin Lomas. Predicting the diversity of internal temperatures from the English residential sector using panel methods. *Applied Energy*, 102:601–621, February 2013.

Mike Kendon, Mark McCarthy, Svetlana Jevrejeva, Andrew Matthews, and Tim Legg. State of the UK climate 2017. *International Journal of Climatology*, 38:135, Jul 2018. ISSN 08998418. doi: 10.1002/joc.5798.

Mike Kendon, Mark McCarthy, Svetlana Jevrejeva, Andrew Matthews, and Tim Legg. State of the UK climate 2018. *International Journal of Climatology*, Jul 2019. doi: 10.1002/joc.6213. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.6213.

H. R Kennard, G.M. Huebner, D. Shipworth, and F. Nicol. Regression Dilution, Bayesian Analysis and Adaptive Thermal Comfort. In *Proceedings of 10th Windsor Conference*, Cumberland Lodge, Windsor Great Park, UK, April 2018. URL http://windsorconference.com/_archive/W18_book_of_abstracts.pdf.

H. R. Kennard, G.M. Huebner, and D. Shipworth. Observational evidence of the seasonal and demographic variation in experienced temperature from 77,743 UK Biobank participants. *Journal of Public Health*, April 2019. doi: 0.1093/pubmed/fdz025.

H.R Kennard, G. M. Huebner, and D. Shipworth. Experienced temperature and health. Technical report, UCL, 2017. URL https://egap.org/registration/5161.

Jacqueline Kerr, Simon Marshall, Suneeta Godbole, Suvi Neukam, Katie Crist, Kari Wasilenko, Shahrokh Golshan, and David Buchner. The relationship between outdoor activity and health in older adults using gps. *International Journal of Environmental Research and Public Health*, 9(12):46154625, Dec 2012. ISSN 1661-7827. doi: 10.3390/ijerph9124615.

Boris R. M. Kingma, Arjan J. H. Frijns, Lisje Schellen, and Wouter van Marken Lichtenbelt. Beyond the classic thermoneutral zone. *Temperature: Multidisciplinary Biomedical Journal*, 1(2):142–149, July 2014. ISSN 2332-8940. doi: 10.4161/temp.29702.

Elselijn Kingma. What is it to be healthy? *Analysis*, 67(294):128–133, April 2007. ISSN 0003-2638. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2239248/.

William Kitchiner. *The art of invigorating and prolonging life, by food, clothes, air, exercise, wine, sleep, &c.* Hurst, Robinson, and Co., London, 1822.

C. R. Kothari. *Research Methodology:Methods and Techniques.* New Age International Pvt Ltd Publishers, 2nd ed. edition edition, Oct 2012. ISBN 978-81-224-3623-5.

David Kriebel, Harvey Checkoway, and Neil Pearce. Exposure and dose modelling in occupational epidemiology. *Occupational and Environmental Medicine*, 64(7):492498, Jul 2007. ISSN 1351-0711. doi: 10.1136/oem.2006.030031.

E.R. Kuras, D.M. Hondula, and J. Brown-Saracino. Heterogeneity in individually experienced temperatures (IETs) within an urban neighborhood: insights from a new approach to measuring heat exposure. *International Journal of Biometeorology*, 59(10): 1363–1372, 2015.

Nicholas L. Lam, Kirk R. Smith, Alison Gauthier, and Michael N. Bates. Kerosene: A review of household uses and their hazards in low and middle-income countries. *Journal of toxicology and environmental health. Part B, Critical reviews*, 15(6):396432, 2012. ISSN 1093-7404. doi: 10.1080/10937404.2012.710134.

Vladimir Ilyich Lenin. *Lenin Collected Works: Volume 37.* Progress Publishers, Moscow, 1977.

B. Levy, A. Ghaem, J. M. Verpillat, and J. P. Martineaud. Effects of ambient and local cutaneous temperatures on hand blood flow. *International Journal of Biometeorology*, 21 (4):341–347, 1977.

Barry Levy and Jonathan A. Patz. *Climate Change and Public Health.* Oxford University Press, Oxford, July 2015. ISBN 978-0-19-020245-3.

Christine Liddell, Chris Morris, Harriet Thomson, and Ciara Guiney. Excess winter deaths in 30 European countries 19802013: a critical review of methods. *Journal of Public Health*, 38(4):806–814, December 2016. ISSN 1741-3842. doi: 10.1093/pubmed/fdv184.

Yuan Luo, Yonghui Zhang, Tao Liu, Shannon Rutherford, Yanjun Xu, Xiaojun Xu, Wei Wu, Jianpeng Xiao, Weilin Zeng, Cordia Chu, and Wenjun Ma. Lagged Effect of Diurnal Temperature Range on Mortality in a Subtropical Megacity of China. *PLOS ONE*, 8(2): e55280, February 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0055280.

A. Macfarlane. Daily mortality and environment in English conurbations. Air pollution, low temperature, and influenza in Greater London. *Journal of Epidemiology & Community Health*, 31(1):54–61, March 1977. ISSN 0143-005X, 1470-2738. doi: 10.1136/jech.31.1.54. URL `https://jech.bmj.com/content/31/1/54`.

Maelstrom. UK biobank: a large-scale prospective epidemiological resource. Technical report, Maelstrom Research, 2019. URL https://www.maelstrom-research.org/mica/individual-study/ukb/.

R. Marin, M. Campos, A. Gomariz, A. Lopez, M. A. Rol, and J. A. Madrid. Complexity Changes in Human Wrist Temperature Circadian Rhythms through Ageing. In José Manuel Ferrández, José Ramón Álvarez Snchez, Félix de la Paz, and F. Javier Toledo, editors, *Foundations on Natural and Artificial Computation*, volume 6686, pages 401–410. Springer Berlin Heidelberg, Berlin, 2011. ISBN 978-3-642-21343-4.

Anil Markandya and Paul Wilkinson. Electricity generation and health. *The Lancet*, 370 (9591):979990, Sep 2007. ISSN 0140-6736, 1474-547X. doi: 10.1016/S0140-6736(07) 61253-7.

Michael Marmot, Ilaria Geddes, Ellen Bloomer, Jessica Allen, and Peter Goldblatt. The Health Impacts of Cold Homes and Fuel Poverty. Technical report, Friends of The Earth. Department of Epidemiology & Public Health University College London, London, 2011.

A Martinez-Nicolas, E Ortiz-Tudela, M. A. Rol, and J. A. Madrid. Uncovering different masking factors on wrist skin temperature rhythm in free-living subjects. *PloS one*, 8(4): e61142, 2013.

A. Mavrogianni, F. Johnson, M. Ucci, A. Marmot, J. Wardle, T. Oreszczyn, and A. Summerfield. Historic Variations in Winter Indoor Domestic Temperatures and Potential Implications for Body Weight Gain. *Indoor + Built Environment*, 22(2):360–375, April 2013.

P McCullagh and J. A. Nelder. *Generalized Linear Models*. Chapman and Hall., London, 1989.

Natalie McGauran, Beate Wieseler, Julia Kreis, Yvonne-Beatrice Schler, Heike Klsch, and Thomas Kaiser. Reporting bias in medical research - a narrative review. *Trials*, 11(37): 15, 2010.

G. R. McGregor, S. Walters, and J. Wordley. Daily hospital respiratory admissions and winter air mass types, Birmingham, UK. *International Journal of Biometeorology*, 43(1): 21–30, July 1999. ISSN 1432-1254. doi: 10.1007/s004840050112.

Mehdi Menai, Vincent T. van Hees, Alexis Elbaz, Mika Kivimaki, Archana Singh-Manoux, and Sverine Sabia. Accelerometer assessed moderate-to-vigorous physical activity and successful ageing: results from the Whitehall II study. *Scientific Reports*, 7, April 2017. ISSN 2045-2322. doi: 10.1038/srep45772.

Christopher W. Meyer, William A. Kimes, and Dean C. Ripple. Determining the thermal response time of temperature sensors embedded in semiconductor wafers. *Measurement Science and Technology*, 19(5):055202, March 2008. ISSN 0957-0233. doi: 10.1088/0957-0233/19/5/055202.

MHCLG. A decent home: definition and guidance, 2006. URL `https://www.gov.uk/government/publications/a-decent-home-definition-and-guidance`. Accessed = 2019-07-24.

MHCLG. English housing survey 2017: stock condition, 2019. URL `https://www.gov.uk/government/statistics/english-housing-survey-2017-stock-condition`. Accessed = 2019-07-24.

Microchip Technology. MCP9700 datasheet, 2007. URL `http://ww1.microchip.com/downloads/en/DeviceDoc/20001942G.pdf`. Accessed = 2017-05-23.

Mohammad Taghi Moghadamnia, Ali Ardalan, Alireza Mesdaghinia, Abbas Keshtkar, Kazem Naddafi, and Mir Saeed Yekaninejad. Ambient temperature and cardiovascular mortality: a systematic review and meta-analysis. *PeerJ*, 5:e3574, Aug 2017. ISSN 2167-8359. doi: 10.7717/peerj.3574.

John William Moore. Art. IX.Notes on Mean Temperature in its Relations to Disease and Mortality, with special reference to the city of Dublin. *Dublin Quarterly Journal of Medical Science*, 48(1):107–122, August 1869. ISSN 0790-2328. doi: 10.1007/BF02944430.

Parker Morris. *Homes for today and tomorrow.* Ministry of Housing and Local Government, London, 1961. URL `http://archive.org/details/op1266209-1001`.

M. Muntean, D. Guizzardi, E. Schaaf, M. Crippa, E. Solazzo, J.G.J Olivier, and E Vignati. *Fossil CO2 emissions of all world countries - 2018 Report.* Publications Office of the European Union, Luxembourg, 2018. ISBN 978-92-79-97240-9.

Makoto Nakayoshi, Manabu Kanda, Rui Shi, and Richard de Dear. Outdoor thermal physiology along human pathways: a study using a wearable measurement system. *International Journal of Biometeorology*, 59(5):503–515, May 2015. ISSN 1432-1254. doi: 10.1007/s00484-014-0864-y.

S. M. Nissilae, Onni Ahola, Ilkka Heikkilae, Antti Ruha, and Harri K. Kopola. Skin temperature profiles on the human chest and in the wrist area. In *Proc. SPIE, Optical and Imaging Techniques for Biomonitoring*, volume 2628, pages 296–303, 1996.

Lennart Nordenfelt. The concepts of health and illness revisited. *Medicine, Health Care and Philosophy*, 10(1):5, September 2006. ISSN 1572-8633. doi: 10.1007/s11019-006-9017-3.

Frank Ocean, Christopher Breaux, Shawn Carter, Tyler Okonma, Adam Feeney, Jacob Ludwig Olofsson, and Rami Dawood. *Biking*. Boys Don't Cry, New Orleans, 2017.

ONS. Deaths occurring by single year of age, by sex, year and month of death, for usual residents of England and Wales: 1993 to 2014. Technical report, Office of National Statistics, 2016. URL https://tinyurl.com/y68t6ugz.

ONS. Excess winter mortality in England and Wales: 2016 to 2017 (provisional) and 2015 to 2016 (final). Technical report, Office of National Statistics, 2017. URL https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/excesswintermortalityinenglandandwales/2016to2017provisionaland2015to2016final.

ONS. Excess winter mortality in england and wales: 2017 to 2018 (provisional) and 2016 to 2017 (final), 2018. URL https://bit.ly/2QpdK9S.

ONS. Census geography, 2019. URL https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography. Accessed = 2019-05-30.

Tadj Oreszczyn, Sung H. Hong, Ian Ridley, and Paul Wilkinson. Determinants of winter indoor temperatures in low income households in England. *Energy and Buildings*, 38(3): 245–252, March 2006.

David Ormandy and Vronique Ezratty. Health and thermal comfort: From WHO guidance to housing strategies. *Energy Policy*, 49:116–121, October 2012. ISSN 0301-4215. doi: 10.1016/j.enpol.2011.09.003.

E. Ortiz-Tudela, A. Martinez-Nicolas, C. Diaz-Mardomingo, S. García-Herranz, I. Pereda-Pérez, Valencia A, Peraita H, Venero C, Madrid JA, and Rol MA. The characterization of biological rhythms in mild cognitive impairment. *BioMed research international*, 2014: 524971, 2014.

Elisabet Ortiz-Tudela, Antonio Martinez-Nicolas, Manuel Campos, Mara Ángeles Rol, and Juan Antonio Madrid. A New Integrated Variable Based on Thermometry, Actimetry and Body Position (TAP) to Evaluate Circadian System Status in Humans. *PLoS Computational Biology*, 6(11), November 2010.

J. Palmer and I. Cooper. United kingdom housing energy fact file, 2013. URL https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/345141/uk_housing_fact_file_2013.pdf. Accessed = 2019-07-24.

Tytti P. Pasanen, Liisa Tyrväinen, and Kalevi M. Korpela. The relationship between perceived health and physical activity indoors, outdoors in built environments, and outdoors in nature. *Applied Psychology: Health and Well-Being*, 6(3):324346, 2014. ISSN 1758-0854. doi: 10.1111/aphw.12031.

Owen Powell and Catatonia. *Don't Need the Sunshine*. Blanco y Negro, London, 1998.

Thomson Prentice. Health, History, and Hard Choices: Funding Dilemmas in a Fast-Changing World. *Nonprofit and Voluntary Sector Quarterly*, 37(1_suppl):63S–75S, March 2008. ISSN 0899-7640. doi: 10.1177/0899764007310533.

SEC Pretlove, T Oreszczyn, I Ridley, T Wilkinson, and D Crowther. A steady-state model for predicting hygrothermal conditions in beds in relation to house dust mite requirements. *Building Services Engineering Research and Technology*, 26(4):301–314, November 2005. ISSN 0143-6244. doi: 10.1191/0143624405bt137oa.

Gary J. Raw, Claire E. Aizlewood, and Ruth M. Hamilton. Building Regulation, health and safety. Technical report, BRE, 2001.

Ole Bjørn Rekdal. Academic urban legends. *Social Studies of Science*, 44(4):638–654, August 2014. ISSN 0306-3127. doi: 10.1177/0306312714535679.

Cambridge Architectural Research. How much energy could be saved by making small changes to every day household behaviours? Technical report, Department of Energy and Climate Change, London, 2012. URL `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/128720/6923-how-much-energy-could-be-saved-by-making-small-cha.pdf`.

N Romeijn and E. J. va Someren. Correlated fluctuations of daytime skin temperature and vigilance. *Journal of Biological Rhythms*, 26(1):68–77, February 2011.

G Rose. Cold weather and ischaemic heart disease. *British Journal of Preventive & Social Medicine*, 20(2):97–100, April 1966. ISSN 0007-1242. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1059032/`.

Ralph L. Rosnow and Robert Rosenthal. Effect sizes for experimenting psychologists. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie exprimentale*, 57(3):221–237, 2003. ISSN 1878-7290(Electronic),1196-1961(Print). doi: 10.1037/h0087427.

Ricardo Forgiarini Rupp, Jungsoo Kim, Richard de Dear, and Enedir Ghisi. Associations of occupant demographics, thermal history and obesity variables with their thermal comfort in air-conditioned and mixed-mode ventilation office buildings. *Building and Environment*, 135:1–9, May 2018. ISSN 0360-1323. doi: 10.1016/j.buildenv.2018.02.049.

Seward B. Rutkove. Effects of temperature on neuromuscular electrophysiology. *Muscle & Nerve*, 24(7):867–882, 2001. ISSN 1097-4598. doi: 10.1002/mus.1084.

Niilo R.I. Ryti, Yuming Guo, and Jouni J.K. Jaakkola. Global Association of Cold Spells and Adverse Health Effects: A Systematic Review and Meta-Analysis. *Environmental Health Perspectives*, 124(1):12–22, January 2016. doi: 10.1289/ehp.1408104.

J. A. Sarabia, M. A. Rol, P. Mendiola, and J. A. Madrid. Circadian rhythm of wrist temperature in normal-living subjects: A candidate of new index of the circadian system. *Physiology & Behavior*, 95(4):570–580, 2008.

Shlomo Sawilowsky. New Effect Size Rules of Thumb. *Journal of Modern Applied Statistical Methods*, 8(2), November 2009. ISSN 1538 - 9472. doi: 10.22237/jmasm/1257035100.

X. Serra-Picamal, R. Roman, J. Escarrabill, A. Garca-Alts, J.M. Argimn, N. Soler, R. Faner, E.M. Carbonell, A. Trilla, and A. Agusti. Hospitalizations due to exacerbations of COPD: A big data perspective. *Respiratory Medicine*, 145:219–225, 2018. doi: 10.1016/j.rmed.2018.01.008.

M. Shilaih, B. M. Goodale, L. Falco, F. Kubler, V. De Clerck, and B. Leeners. Modern fertility awareness methods: Wrist wearables capture the changes of temperature associated with the menstrual cycle. *Bioscience reports*, November 2017.

S. Y. Sim, M . J. Koh, K. M. Joo, S. Noh, S. Park, Y. H. Kim, and K. S. Park. Estimation of Thermal Sensation Based on Wrist Skin Temperatures. *Sensors (Basel, Switzerland)*, 16(4):420, March 2016.

Adam Smith. *The Wealth of Nations*. W. Strahan and T. Cadell, London, 1776.

Philip L. Staddon, Hugh E. Montgomery, and Michael H. Depledge. Climate warming will not decrease winter mortality. *Nature Climate Change*, 4(3):190–194, March 2014. ISSN 1758-678X. doi: 10.1038/nclimate2121.

Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, Bette Liu, Paul Matthews, Giok Ong, Jill Pell, Alan Silman, Alan Young, Tim Sprosen, Tim Peakman, and Rory Collins. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Medicine*, 12(3), March 2015. ISSN 1549-1277. doi: 10.1371/journal.pmed.1001779.

Grace Sum, Yun Ru Tan, Song-Iee Hong, and Gerald Choon-Huat Koh. Psychosocial and physical factors associated with depression in older adults living in studio apartments: A cross-sectional study of a multi-ethnic asian population. *Proceedings of Singapore Healthcare*, Jul 2019. ISSN 2010-1058. doi: 10.1177/2010105819860853.

Shengzhi Sun, Wangnan Cao, Tonya G. Mason, Jinjun Ran, Hong Qiu, Jinhui Li, Yang Yang, Hualiang Lin, and Linwei Tian. Increased susceptibility to heat for respiratory hospitalizations in hong kong. *Science of The Total Environment*, 666:197204, May 2019. ISSN 0048-9697. doi: 10.1016/j.scitotenv.2019.02.229.

C Szmaragd and G Leckie. Introduction to multilevel modelling - r practical. lemma vle module 5, 1-42. `http://www.bristol.ac.uk/cmm/learning/online-course`, 2011. Accessed: 2018-03-30.

Toshiyo Tamura, Ming Huang, and Tatsuo Togawa. Current Developments in Wearable Thermometers. *Advanced Biomedical Engineering*, 7(0):88–99, 2018. ISSN 2187-5219. doi: 10.14326/abe.7.88. URL `https://www.jstage.jst.go.jp/article/abe/7/0/7_7_88/_article`.

Louise M Tanner, Suzanne Moffatt, Eugene M G Milne, Susanna D H Mills, and Martin White. Socioeconomic and behavioural risk factors for adverse winter health and social outcomes in economically developed countries: a systematic review of quantitative observational studies. *Journal of Epidemiology and Community Health*, 67(12):1061–1067, December 2013. ISSN 0143-005X, 1470-2738. doi: 10.1136/jech-2013-202693.

Mike Tipton. Humans: A homeothermic animal that needs perturbation? *Experimental Physiology*, 104(1):12, 2019. ISSN 1469-445X. doi: 10.1113/EP087450.

UK Biobank. Material transfer agreement, 2017. URL `http://www.ukbiobank.ac.uk/wp-content/uploads/2012/09/Material-Transfer-Agreement.pdf`. Accessed = 2019-05-30.

UK Biobank. Data-field 0-192, 2019. URL `http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=90192`. Accessed = 2017-01-01.

UK Government. Climate Change Act 2008, 2008. URL `http://www.legislation.gov.uk/ukpga/2008/27/contents`.

K. Vadodaria, D. L. Loveday, and V. Haines. Measured winter and spring-time indoor temperatures in UK homes over the period 19692010: A review and synthesis. *Energy Policy*, 64:252–262, January 2014.

Wouter Marken van Lichtenbelt, Mark Hanssen, Hannah Pallubinsky, Boris Kingma, and Lisje Schellen. Healthy excursions outside the thermal comfort zone. *Building Research & Information*, 45(7):819–827, October 2017. ISSN 0961-3218. doi: 10.1080/09613218.2017.1307647.

C. H. Vinkers, R. Penning, J. Hellhammer, J. C. Verster, J. H. G. M Klaessens, B. Olivier, and C. J. Kalkman. The effect of stress on core and peripheral body temperature in humans. *Stress*, 16(5):520–530, 2013.

Daniela Wey, Andreas Bohn, and Luiz Menna-Barreto. Daily rhythms of native Brazilians in summer and winter. *Physiology & Behavior*, 105(3):613–620, February 2012.

Matthew D. White, Catharine M. Bosio, Barry N. Duplantis, and Francis E. Nano. Human body temperature and new approaches to constructing temperature-sensitive bacterial vaccines. *Cellular and molecular life sciences: CMLS*, 68(18):3019–3031, September 2011. ISSN 1420-9071. doi: 10.1007/s00018-011-0734-2.

P. Wilkinson, M. Landon, B. Armstrong, S. Stevenson, S. Pattenden, M. McKee, and T. Fletch. Cold comfort: The social and environmental determinants of excess winter deaths in England, 1986-1996. Technical report, The Policy Press, Bristol, 2001. URL `https://tinyurl.com/y2r884lv`.

Paul Wilkinson, Kirk R Smith, Sean Beevers, Cathryn Tonne, and Tadj Oreszczyn. Energy, energy efficiency, and the built environment. *The Lancet*, 370(9593):1175–1187, September 2007. ISSN 01406736. doi: 10.1016/S0140-6736(07)61255-0. URL `https://linkinghub.elsevier.com/retrieve/pii/S0140673607612550`.

Bill Withers. *Harlem*. Sussex, Hollywood, 1971.

Rachel Wookey, Angie Bone, Catriona Carmichael, and Anna Crossley. Minimum home temperature thresholds for health in winter A systematic literature review. Technical report, Public Health England, 2014. URL `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/468196/Min_temp_threshold_for_homes_in_winter.pdf`.

Z. Wu, N. Li, H. Cui, J. Peng, H. Chen, and P. Liu. Using Upper Extremity Skin Temperatures to Assess Thermal Comfort in Office Buildings in Changsha, China. *International journal of environmental research and public health*, 14(10), September 2017.

Yu-Fei Xing, Yue-Hua Xu, Min-Hua Shi, and Yi-Xin Lian. The impact of PM2.5 on the human respiratory system. *Journal of Thoracic Disease*, 8(1):E69E74, Jan 2016. ISSN 2072-1439. doi: 10.3978/j.issn.2072-1439.2016.01.19.

Zhiwei Xu, Ruth A. Etzel, Hong Su, Cunrui Huang, Yuming Guo, and Shilu Tong. Impact of ambient temperature on children's health: A systematic review. *Environmental Research*, 117:120–131, August 2012. ISSN 0013-9351. doi: 10.1016/j.envres.2012.07.002.

Ye Xiaofang, Wolff Rodney, Yu Weiwei, Vaneckova Pavla, Pan Xiaochuan, and Tong Shilu. Ambient Temperature and Morbidity: A Review of Epidemiological Evidence. *Environmental Health Perspectives*, 120(1):19–28, January 2012. doi: 10.1289/ehp.1003198.

Yunquan Zhang, Minjin Peng, Lu Wang, and Chuanhua Yu. Association of diurnal temperature range with daily mortality in england and wales: A nationwide time-series study. *Science of the Total Environment*, 619620:291300, Apr 2018. ISSN 0048-9697. doi: 10.1016/j.scitotenv.2017.11.056.