

# BMJ Open

## Taxonomy of Self-reported Sedentary behaviour Tools (TASST): a framework for development, comparison and evaluation

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2016-013844
Article Type:	Research
Date Submitted by the Author:	10-Aug-2016
Complete List of Authors:	Dall, Philippa; Glasgow Caledonian University School of Health and Life Sciences, Department of Psychology, Social Work & Allied Health Sciences Coulter, Elaine; University of Glasgow Fitzsimons, Claire F.; Univ Edinburgh, ISPEHS Skelton, Dawn; Glasgow Caledonian University, School of Health & Life Sciences Chastin, Sebastian; Glasgow Caledonian University
<b>Primary Subject Heading</b>:	Public health
Secondary Subject Heading:	Epidemiology, Research methods, Sports and exercise medicine
Keywords:	sedentary behaviour, sitting, population surveillance, measurement, validation

SCHOLARONE™  
Manuscripts

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1 **Title:** TAxonomy of Self-reported Sedentary behaviour Tools (TASST): a framework for development,  
2 comparison and evaluation

3

4 **Authors:**

5 Dall PM<sup>1</sup>, Coulter EH<sup>2</sup>, Fitzsimons CF<sup>3</sup>, Skelton DA<sup>1</sup>, Chastin SFM<sup>1</sup>, on behalf of the Seniors USP Team.

6

7 <sup>1</sup> Institute for Applied Health Research, Glasgow Caledonian University, Cowcaddens Road, Glasgow,  
8 G4 0BA

9 <sup>2</sup> Department of Nursing and Healthcare, Glasgow University, 59 Oakfield Avenue, Glasgow, G12 8LL

10 <sup>3</sup> Sport, Physical Education and Health Sciences Institute, Edinburgh University, St Leonard's Land,  
11 Holyrood Road, Edinburgh, EH8 8AQ

12

13 corresponding author:

14 Dr Philippa Dall

15 School of Health & Life Sciences

16 Glasgow Caledonian University

17 Cowcaddens Road

18 Glasgow

19 G4 0BA

20 philippa.dall@gcu.ac.uk

21

22 Word count: 4,746

23

1  
2  
3 24 **ABSTRACT**  
4  
5

6 25 **Objective:** Sedentary behaviour (SB) has distinct deleterious health outcomes, yet there is no  
7  
8 26 consensus on best practice for measurement. This study aimed to identify the optimal tool for  
9  
10 27 population surveillance of SB, using a systematic framework.  
11

12  
13 28 **Design:** A framework, Taxonomy of Self-report SB Tools (TASST), was developed based on a  
14  
15 29 systematic inventory of existing tools. The inventory was achieved through a systematic review of  
16  
17 30 studies reporting SB and tracing back to the original description. A systematic review of the accuracy  
18  
19 31 and sensitivity to change of these tools was then mapped against TASST domains.  
20  
21

22 32 **Data Sources:** Systematic searches were conducted via EBSCO.  
23  
24

25 33 **Eligibility Criteria for selecting studies:** The inventory included tools measuring SB in adults that  
26  
27 34 could be self-completed at one sitting, and excluded tools measuring SB in specific populations or  
28  
29 35 contexts. The systematic review included studies reporting on the accuracy against an objective  
30  
31 36 measure of SB and/or sensitivity to change of a tool in the inventory.  
32  
33

34  
35 37 **Results:** The systematic review identified 32 distinct tools comprising 141 questions. The TASST  
36  
37 38 framework has four domains (type of assessment, recall period, temporal unit, and assessment  
38  
39 39 period), which characterised all self-report SB tools. Fourteen studies evaluated accuracy and/or  
40  
41 40 sensitivity to change representing only 6 taxa. Assessing SB as a sum of behaviours and using a  
42  
43 41 previous day recall were the most promising features of existing tools. Accuracy was poor for all  
44  
45 42 existing tools, with both under and over estimation of SB. There was a complete lack of evidence  
46  
47 43 about sensitivity to change.  
48  
49

50  
51 44 **Conclusions:** Despite the limited evidence, mapping existing SB tools onto the TASST framework has  
52  
53 45 enabled informed recommendations to be made about the most promising features for a  
54  
55 46 surveillance tool, and to identify the aspects on which future research and development of SB  
56  
57 47 surveillance tools should focus.  
58  
59  
60

1  
2  
3 48 **Systematic Review Registration**  
4  
5

6 49 PROSPERO (CRD42014009851)  
7  
8

9 50  
10

11 51 **KEY WORDS:**  
12

13  
14  
15 52 sedentary behaviour; sitting; population surveillance; measurement; validation  
16  
17

18 53  
19

20  
21 54 **STRENGTHS AND LIMITATIONS OF THIS STUDY:**  
22

- 23  
24 55 • A systematic approach was taken towards classifying self-reported measures of sedentary  
25  
26 56 behaviour, allowing a structured approach to measurement in the future  
27  
28 57 • An example of use of the framework is presented, mapping accuracy and sensitivity to  
29  
30 58 change of self-report sedentary behaviour measures on to the framework  
31  
32 59 • Although designed to be generic, the TASST framework was developed excluding tools  
33  
34 60 measuring sedentary behaviour in specialised contexts, e.g. children, and the framework  
35  
36 61 may therefore not cover some aspects of these tools  
37  
38 62 • There is the potential for a language bias, as full-text articles not in English were not  
39  
40 63 included in the systematic reviews.  
41  
42  
43  
44 64

1  
2  
3 65 **BACKGROUND**  
4  
5

6 66 Physical inactivity is currently at pandemic levels [1] and is a global public health concern. Sedentary  
7  
8 67 behaviour (SB), an umbrella term for all waking time spent in non-exercising sitting or reclining  
9  
10 68 postures [2, 3] such as sitting during work, motorised transport or watching TV, is the largest  
11  
12 69 contributor to inactivity [4,5]. Higher levels of SB have been associated with poor physical and  
13  
14 70 mental health, increased risk of chronic disease and less successful ageing [6-9]. Consequently,  
15  
16 71 several countries, including the UK, have issued recommendations to reduce SB at all ages as part of  
17  
18 72 their national physical activity guidelines [10]. Population surveillance is urgently needed to monitor  
19  
20 73 the impact of such policy, track changes in SB over time, and to evaluate public health interventions  
21  
22 74 targeting SB. In order to provide effective surveillance upon which to base future policy decisions,  
23  
24 75 such surveillance tools should be accurate (provide a true measure of the actual amount of SB in a  
25  
26 76 population) and sensitive to change (provide the true difference in SB between two measurement  
27  
28 77 time points) [11].  
29  
30  
31  
32  
33  
34  
35

36 79 Objective body worn sensors, that measure posture, demonstrate good accuracy for measuring total  
37  
38 80 duration of SB against the gold standard of direct observation [12], but they are expensive and  
39  
40 81 challenging to use for population surveillance. Self-report tools provide a pragmatic choice for  
41  
42 82 population surveillance and have the potential to provide context rich information, useful for  
43  
44 83 intervention development [13]. To date, surveys assessing SB have predominantly used self-report  
45  
46 84 tools [14], which are generally adapted from tools not specifically designed to measure that  
47  
48 85 behaviour (e.g. tools designed to measure physical activity) [15], and which have not been evaluated  
49  
50 86 for population surveillance purposes [14]. No framework currently exists with which to describe and  
51  
52 87 compare SB self-report tools, meaning there is currently no way of systematically selecting an  
53  
54 88 appropriate tool. A previous systematic review of the measurement characteristics of self-report  
55  
56 89 tools measuring SB, reported acceptable to good reliability but low to moderate correlation with a  
57  
58  
59  
60

1  
2  
3 90 (non-gold standard) criterion measure [13]. This suggests that self-report measures of SB are  
4  
5 91 acceptable tools to establish epidemiological evidence of an association between SB and health [13].  
6  
7 92 However, it is possible that the scale of the problem may be vastly underestimated, as differences of  
8  
9 93 2-4 hours per day (approximately 20% of SB) have been reported between self-report and objective  
10  
11 94 tools [16].  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33

34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

95  
96 The primary aim of this study was to identify, in a systematic manner, the optimal tool to measure  
97 SB for use in population surveillance. To fulfil the primary aim, a framework was created to describe  
98 the features of self-report tools measuring SB, the TAXonomy of Self-report Sedentary behaviour  
99 Tools (TASST). A systematic inventory of existing self-report tools to measure SB was mapped onto  
100 TASST, and the measurement characteristics of these tools, focussing on accuracy and sensitivity to  
101 change, were evaluated, with explicit reference to the domains of the taxonomy framework.

## 102 103 **METHODS**

104 The study protocol (PROSPERO CRD42014009851), was conducted in three phases. In phase 1 an  
105 exhaustive inventory of self-reported tools to measure SB in adults and older adults was established  
106 using a structured search protocol. Phase 2 was the development of a taxonomy based on content  
107 analysis of the items and questions in the tools. In phase 3, a systematic literature review of the  
108 measurement characteristics of the tools in the inventory was conducted and mapped onto the  
109 taxonomy.

### 110 111 **Phase 1: Systematic inventory of self-report tools**

1  
2  
3 112 The aim of the systematic inventory was to compile an exhaustive list of self-report tools which  
4  
5 113 could be used to measure SB in adults and older adults. Since the aim was to identify tools and not  
6  
7 114 to identify articles, this stage does not have the same methodology as a systematic literature review.  
8  
9 115 A literature search was conducted in October 2013, for articles reporting SB as an outcome measure.  
10  
11 116 From this review, an list of self-report tools which measured SB was compiled. References lists were  
12  
13 117 reviewed and experts consulted to identify any additional tools to include in the inventory. The  
14  
15 118 inventory then was consolidated to amalgamate tools referred to by different names, and to trace  
16  
17 119 back to the original version. Articles which added significant new questions to tools were included  
18  
19 120 as a separate tool. Tools used in a single study and those without names/acronyms were included as  
20  
21 121 separate tools.  
22  
23  
24  
25  
26  
27

28 123 To be included in the inventory, tools had to: be suitable for use for large scale population studies of  
29  
30 124 adults or older adults; measure SB or a proxy measure of SB; and be suitable for self-completion by  
31  
32 125 the respondent at a single point in time. Tools were excluded from the inventory: if they were  
33  
34 126 designed specifically to assess SB in children or other specialised populations (e.g. medical  
35  
36 127 conditions); if they were designed specifically to assess SB in a specialised context, (e.g. workplace or  
37  
38 128 care settings); if continuous reporting over extended periods of time was required (e.g. diaries or  
39  
40 129 time-use surveys); or if significant interviewer interactions were required. Self-report tools that  
41  
42 130 could be administered by telephone or interview were not automatically excluded, however tools  
43  
44 131 such as the PDR (Previous Day Recall) [17], in which the interviewer works through lists of several  
45  
46 132 hundred items, were excluded.  
47  
48  
49

50  
51 133

52  
53  
54 134 **Phase 2: Development of a taxonomy**  
55  
56  
57  
58  
59  
60

1  
2  
3 135 The original text was extracted for each question relating to SB in each of the self-report tools  
4  
5 136 identified in the inventory. Content analysis was conducted on the text to extract all of the  
6  
7 137 attributes in the questions that were used to describe and constrain what aspect of SB was  
8  
9 138 measured by that question. For example, in the question “During the last 7 days, how much time did  
10  
11 139 you usually spend sitting on a week day?”, attributes extracted relating to the measurement of SB  
12  
13 140 would be “during the last 7 days”, “time spent sitting” and “on a week day”. Attributes were then  
14  
15 141 grouped into mutually exclusive domains covering similar aspects of measurement, and categories  
16  
17 142 within those domains were defined iteratively. A new category was created each time a tool did not  
18  
19 143 fit within an existing category. The full taxonomy was then assembled and streamlined by merging  
20  
21 144 categories with overlapping meaning. Finally, consideration was given to potential future  
22  
23 145 developments of self-report tools to measure SB, by adding any categories to the taxonomy  
24  
25 146 considered useful in the future. The resulting taxonomy was then tested by ensuring that all tools  
26  
27 147 could be classified similarly by two independent researchers and that the taxonomy fully defined the  
28  
29 148 tool.  
30  
31  
32  
33  
34  
35  
36

### 37 150 **Phase 3: Systematic review of measurement characteristics**

38  
39  
40 151 Finally, a systematic literature search in relevant health databases was conducted in December 2014  
41  
42 152 via EBSCO host. The search combined the name of the tool including variants and acronyms (except  
43  
44 153 where the acronym was also a common word, e.g. PAST, MOST), with search terms relating to  
45  
46 154 measurement characteristics (valid\* /reliab\* /repons\* /sensitiv\* /calibrat\* /accura\* /agreement  
47  
48 155 /psychometric\* /clinimetric\* /“measurement characteristics” /Reliability and Validity (MeSH)).  
49  
50 156 Articles were included only if they reported in English on the accuracy of a tools in the inventory  
51  
52 157 against an objective criterion measure of SB, and/or sensitivity to change..  
53  
54  
55

56 158  
57  
58  
59  
60



1  
2  
3 159 Exclusion by title, then abstract, then full-text was conducted independently by two researchers  
4  
5 160 from a pool of four [PD, EC, CF, SC]. In the case of disagreement, the article was carried forward in  
6  
7 161 to the next round, or at full-text stage a third researcher was consulted to ensure consensus. Data  
8  
9 162 (tool, criterion, population, statistical analysis, accuracy of sedentary behaviour, sensitivity to change  
10  
11 163 of sedentary behaviour) was extracted and quality was assessed independently by two researchers  
12  
13 164 from a pool of three [PD, CF, SC]. Disagreements were resolved by discussion. Quality was assessed  
14  
15 165 using QUALSYST [18], modified to include an additional item for the criterion measure. As per the  
16  
17 166 QUALSYST guidelines, the quality score for the article (range 0-1) was used to identify common  
18  
19 167 methodological strengths and flaws, rather than as an objective representation of high/low quality.  
20  
21 168 Accuracy and sensitivity to change extracted from included articles were reported for tools in  
22  
23 169 relation to the TASST taxonomy.  
24  
25  
26  
27  
28  
29  
30  
31

170

## 171 **RESULTS**

### 172 **Inventory**

32  
33  
34  
35  
36 173 The systematic inventory identified 32 distinct self-report tools used to measure SB in adults and  
37  
38 174 older adults (Table 1). The International Physical Activity Questionnaire (IPAQ) has four different  
39  
40 175 versions included in the inventory (combinations of the long and short versions, and last seven days  
41  
42 176 and usual week recall). The 45 and Up study asked different questions in its baseline and follow-up  
43  
44 177 questionnaires, which have been included as separate tools. Three tools, termed “modified”  
45  
46 178 versions, were included where questions had been added or modified to the original tool (EPAQ2,  
47  
48 179 IPAQ-L, NHANES), and were considered to form a substantially different version. Some tools  
49  
50 180 identified were used in only a single study, and these were included in the inventory, referred to by  
51  
52 181 the study name. These 32 tools comprised of 141 individual questions, consisting of between 1 and  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

182 20 questions per tool. An evaluation of the content of these individual items formed the basis of the

183 TASST taxonomy.

184

For peer review only

185 Table 1 Tools measuring SB for population surveillance identified in the inventory

Acronym	Name of Tool/Study	Key reference
45Up-B	45 and Up study, baseline questionnaire	[19]
45Up-F	45 and Up study, follow up questionnaire	[19]
ACS2	American Cancer Society, Cancer Prevention Study cohort II	[20]
ALTS	Australian Leisure Time Sitting questionnaire	[21]
AusDiab	The Australian Diabetes Obesity and Lifestyle study	[22]
CCHS	Canadian Community Health Survey	[23]
CFS	Canadian Fitness Survey	[24]
CHAMPS	Community Health Activities Model Program for Seniors physical activity questionnaire	[15]
ELSA	English Longitudinal Study of Ageing	[25]
EPAQ2	European Prospective Investigation of Cancer (EPIC)-Norfolk Physical Activity Questionnaire	[26]
mod EQPAQ2	modified version of the EPIC-Norfolk Physical Activity Questionnaire	[27]
GPAQ	Global Physical Activity Questionnaire	[28]
HSE	Health Survey for England	[29]
HUNT3	Nord-Trøndelag Health Study 3	[30]
IPAQ-L I7d	International Physical Activity Questionnaire, Long version, last 7 days	[31]
IPAQ-L uw	International Physical Activity Questionnaire, Long version, usual week	[31]
mod IPAQ-L	modified version of the International Physical Activity Questionnaire, Long version	[32]
IPAQ-S I7d	International Physical Activity Questionnaire, Short version, last 7 days	[31]
IPAQ-S uw	International Physical Activity Questionnaire, Short version, usual week	[31]
LASA	Longitudinal Aging Study Amsterdam	[33]
MLTPAQ	Minnesota Leisure Time Physical Activity Questionnaire	[34]
MOST	Measuring Older adults' Sedentary Time questionnaire	[35]
NHANES	National Health and Nutrition Examination Survey	[36]
mod NHANES	modified version of the National Health and Nutrition Examination Survey	[37]
NHS2	Nurses Health Survey II	[38]
NIH-AARP DHS	National Institutes of Health – American Association of Retired Persons (NIH-AARP) Diet and Health Survey	[39]
NSWPAS	New South Wales Physical Activity Survey	[40]
PASE	Physical Activity Scale for the Elderly	[41]
PAST	Past-day Adults Sedentary Time questionnaire	[42]
PCSpa	prospective cohort study (Spain)	[43]
SBQ	Sedentary Behaviour Questionnaire	[44]
SHS	Scottish Health Survey	[45]

186 Acronym: the commonly used acronym of the tool, or the short identifier adopted for this article Name of Tool: either the  
 187 name of the tool, or the name of the single study using these questions/tool. Key reference: references provided here are  
 188 not exhaustive, but refer either to an early or well cited description of the tool, or the study in which the tool was used or  
 189 adapted.

1  
2  
3 190 **T**Axonomy for Self-report Sedentary behaviour Tools (TASST)  
4  
5

6 191 The taxonomy derived from the inventory of self-report tools to measure SB (Figure 1) comprises of  
7  
8 192 four domains, which characterise different aspects of the tool: type of assessment, recall period,  
9  
10 193 temporal unit, and assessment period. All four aspects are required to describe the tool. Within  
11  
12 194 each aspect, the taxonomy functions as a tree, meaning you can identify a single end point (taxon)  
13  
14 195 which fully describes each question in a tool..  
15  
16

17  
18 196

19  
20 197 The type of assessment domain of the taxonomy covers the way that the outcome of time spent in  
21  
22 198 SB is derived from the tool. Tools can either ask about a single aspect of SB (1.1 single item), or a  
23  
24 199 composite aspect (1.2 composite). Tools using a single item of assessment will generate all of their  
25  
26 200 information about SB within the relevant period of assessment in a single question. That single item  
27  
28 201 can either ask about sitting time directly (1.1.1 direct measure) or it can ask about a single behaviour  
29  
30 202 related to SB which is then used as a proxy measure of SB duration (1.1.2 proxy measure).  
31  
32

33 203 Composite items of assessment ask multiple questions about several aspects of SB for the same  
34  
35 204 period of assessment. One form of composite item would be to ask about the pattern (i.e.  
36  
37 205 frequency and timing) of SB accumulated throughout the recall period (1.2.1 pattern). However, the  
38  
39 206 most common form of composite item is created as a sum (1.2.2 sum) of the time spent in SB in a  
40  
41 207 range of different activities or situations. The sum can either be formed from questions asking about  
42  
43 208 specific behaviours (1.2.2.1), activities such as TV viewing, hobbies, talking with friends, or they can  
44  
45 209 be based on domains (1.2.2.2), locations or situations where you can sit, such as at home, for  
46  
47 210 transport and at work.  
48  
49

50  
51 211

52  
53  
54 212 The recall period is total time over which the respondent is asked to consider their SB when  
55  
56 213 answering the questions. The recall period can be anchored to the present time in which case it  
57  
58  
59  
60

1  
2  
3 214 refers to a specific length of time prior to now, for example yesterday (2.1 previous day), last week  
4  
5 215 (2.2 previous week), or a longer period such as the last month or year (2.3 longer). The recall period  
6  
7 216 can also be unanchored (2.4), in which case the respondent is not asked about a specific period but  
8  
9 217 is asked about a general period of time, for example asking about SB in a typical week.  
10

11  
12 218

13  
14  
15 219 The temporal unit is the duration within the recall period that a respondent is asked to report their  
16  
17 220 SB for. For example, in the question “on a typical day last week, how long did you sit?” the recall  
18  
19 221 period is the previous week, but the temporal unit is a day. Within the taxonomy, the temporal  
20  
21 222 units may be a day (3.1), a week (3.2) or longer (3.3). Within a particular recall period, it is possible  
22  
23 223 to have any temporal unit that is of identical or shorter duration than the recall period.  
24  
25  
26

27 224

28  
29  
30 225 The period of assessment is completed by identifying any specific restrictions that are placed on the  
31  
32 226 type of temporal unit recalled. The categories within the assessment period domain clarify whether  
33  
34 227 a respondent is asked questions regarding a particular type of day, for example only about week  
35  
36 228 days (4.1), only weekend days (4.2), or is asked about weekdays and weekend days in separate  
37  
38 229 questions (4.3 both). Additionally, the assessment period domain can identify if a respondent is  
39  
40 230 asked about particular sub divisions of the day (4.4) in separate questions, for example time spent  
41  
42 231 sitting before 6pm. The final taxon in the assessment period is termed ‘not defined’ (4.5), this  
43  
44 232 represents the situation where a respondent is asked about all temporal units (e.g. days) within the  
45  
46 233 recall period (e.g. last week) without any specific distinction being made between them. It is a  
47  
48 234 global category, which usually represents a decision not to separate out these categories, as  
49  
50 235 opposed to a failure to define this domain.  
51  
52

53 236

54  
55  
56  
57 237 **Mapping the Inventory on to the Taxonomy**  
58  
59  
60

1  
2  
3 238 The 32 tools identified in the inventory were mapped against the TASST taxonomy (Table 2). Over  
4  
5 239 half of the tools in the inventory (n=17) used a single item of assessment, thirteen used a direct  
6  
7 240 measure and seven used a proxy measure. Three tools asked single item questions about both a  
8  
9 241 direct measure and a proxy measure, but not in a manner in which they could be used as a sum, and  
10  
11 242 have therefore been included in the count for both taxa. Proxy measures were predominantly based  
12  
13 243 on TV viewing (n=5). Fifteen tools used composite assessment, all of which used a sum as that  
14  
15 244 composite item. The vast majority of sums were formed from questions asking about different  
16  
17 245 behaviours (n=14), with only one sum formed from questions asking about different domains. The  
18  
19 246 tools using a sum of behaviours generally included the common proxy measures of TV viewing  
20  
21 247 (n=14) and computer use (n=12) within the sum. Many tools included questions for behaviours  
22  
23 248 based on leisure pursuits (n=9), in social contexts (n=6), and during transportation (n=8). Often  
24  
25 249 several behaviours of each type were considered in separate questions (e.g. asking about time  
26  
27 250 sitting while reading separately from time spent sitting listening to music). Questions based on time  
28  
29 251 working were included in five tools, but were explicitly excluded in four tools. Less frequently, tools  
30  
31 252 included questions based on rest (n=2), or used an "other" category to cover circumstances not  
32  
33 253 explicit within the questions (n=3).  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Table 2: Mapping of the tools measuring SB identified in the inventory onto the TASST taxonomy.

Taxonomy Item	N	Tools	Accuracy	Sensitivity to change
<b>1</b>				
<b>Type of Assessment</b>				
1.1	17		<b>Underestimate with large systematic and a random error</b>	+
1.1.1	13	45Up-B; ACS2; AusDiab; CFS; GPAQ; HUNT3; IPAQ-L I7d; IPAQ-L uw; IPAQ-S I7d; IPAQ-S uw; NIH-AARP DHS; PASE; PCSpa		
1.1.2	7	45Up-B; AusDiab; ELSA; MLTPAQ; NIH-AARP DHS; NSWPAS; SHS		
1.2	15		<b>Smaller systematic error but there is a potential to overestimate</b>	+
1.2.1	0			
1.2.2	15			
1.2.2.1	14	45Up-F; ALTS; CCHS; CHAMPS; EPAQ2; mod EPAQ2; HSE; mod IPAQ-L; LASA; MOST; NHANES; mod NHANES; PAST; SBQ;		
1.2.2.2	1	NHS2		
<b>2</b>				
<b>Recall period</b>				
2.1	1	PAST	<b>+</b>	-
2.2	8	45Up-F; ALTS; AusDiab; IPAQ-L I7d; IPAQ-S I7d; mod IPAQ-L; MOST; PASE	-	+
2.3	8	ACS2; CCHS; CHAMPS; EPAQ2; mod EPAQ2; HSE; NHANES; NIH-AARP DHS	-	-
2.4	15	45Up-B; CFS; ELSA; GPAQ; HUNT3; IPAQ-L uw; IPAQ-S uw; LASA; MLTPAQ; mod NHANES; NHS2; NSWPAS; PCSpa; SBQ; SHS		
<b>3</b>				
<b>Temporal Unit</b>				
3.1	27	45Up-B; 45Up-F; ACS2; AusDiab; CFS; ELSA; EPAQ2; mod EPAQ2; GPAQ; HSE; HUNT3; IPAQ-L I7d; IPAQ-L uw; IPAQ-S I7d; IPAQ-S uw; mod IPAQ-L; LASA; MLTPAQ; NHANES; mod NHANES; NIH-AARP DHS; NSWPAS; PASE; PAST; PCSpa; SBQ; SHS	<b>+</b>	+
3.2	5	ALTS; CCHS; CHAMPS; MOST; NHS2	-	-
3.3	0		-	-
<b>4</b>				
<b>Assessment Period</b>				
4.1	2	IPAQ-S I7d; IPAQ-S uw	-	+
4.2	0		-	-
4.3	12	45Up-F; AusDiab; ELSA; HSE; IPAQ-L I7d; IPAQ-L uw; mod IPAQ-L; LASA; NSWPAS; PCSpa; SBQ; SHS	<b>+</b>	-
4.4	1	EPAQ2	<b>+</b>	-
4.5	18	45Up-B; ACS2; ALTS; CCHS; CFS; CHAMPS; EPAQ2; mod EPAQ2; GPAQ; HUNT3; MLTPAQ; MOST; NHS2; NHANES; mod NHANES; NIH-AARP DHS; PASE; PAST	<b>Better for older adults</b>	+

Full names for the acronyms reported in the Tools column can be found in Table 1. Recommendations in bold are backed by evidence from the systematic review. Recommendations which are not bold are

theoretical but no evidence could be found in the literature. + represents a positive attribute; - a negative attribute.

1  
2  
3 257 About half the tools in the inventory used an unanchored recall period (n=15), eight used a previous  
4  
5 258 week recall period, and eight used a longer recall period. Only a single tool in the inventory used a  
6  
7 259 previous day recall period. The majority of tools used a temporal unit of a day (n=27), with five using  
8  
9 260 a temporal unit of a week. A single question within the EPAQ2 questionnaire was based on a  
10  
11 261 temporal unit longer than a week, but the other three questions in that tool were based on a  
12  
13 262 temporal unit of a day. Just over half the tools (n=18) did not define specific days or time periods in  
14  
15 263 their questions, but asked about the temporal unit within the recall period as a single entity.  
16  
17  
18 264 Fourteen tools used questions specifically referring to week or weekend days, twelve asking about  
19  
20 265 both week and weekend days, while two asked only about week days. Only one tool referred to  
21  
22 266 specific sub-divisions of the day in their questions.  
23  
24  
25  
26  
27

#### 28 **Systematic search for measurement characteristics**

29  
30  
31 269 The systematic search returned 5,640 references, and after removal of duplicate and assessment  
32  
33 270 against exclusion criteria, a total of 14 studies were included in the review (figure 2, table 3).  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



271 Table 3: Measurement characteristics of tools measuring SB, presented by tool and taxon

Tool	Taxon (refer to figure 1)	N	Population (Country)	Criterion measure (definition of SB)	QUALSYST Score	Agreement (hours/day) tool - criterion [limit of agreement]	Sensitivity to change	Ref
IPAQ-Long l7d	1.1.1/2.2/3.1/4.3	1508	A & OA (Greenland)	actiHeart (<1.5MET)	0.67	-3.0 [not reported] for adults -6.0 [not reported] for older adults	--	[46]
		542	A (Netherlands)	Actigraph (<100 count/min)	0.78	-1.6 [-6.4 3.2]	--	[47]
		980	A (Sweden)	Actigraph (<100 count/min)	0.67	+2.2 [-4.5 9.5]	--	[48]
		69	A (UK)	activPAL (sitting/lying postures)	0.78	-2.2 [-7.22 3.71]	--	[16]
		317	A (Chile)	Actigraph (<100 count/min)	0.78	-1.1 [-3.8 1.5]	--	[49]
IPAQ-Short l7d	1.1.1/2.2/3.1/4.1	1751	A & OA (Norway)	Actigraph (<100 count/min)	0.67	-1.8 [not reported] for adults +3.5 [not reported] for older adults	--	[50]
		144	A (Nigeria)	Actigraph (<100 count/min)	0.78	-3.0 [-8.5 2.5]	--	[51]
		54	OA (Sweden)	Actigraph (<100 count/min)	0.56	-1.5 [not reported]	--	[52]
		127	OA (USA)	Actigraph (<50 count/min)	0.72	-4.4 [-10.0 -1.4]	--	[53]
CHAMPS	1.2.2.1/2.3/3.2/4.5	870	OA (USA)	Actigraph (<100 count/min)	0.72	-6.8 [-10.6 2.4]	--	[15]
		58	OA (USA)	Actigraph (<100 count/min)	0.72	-5.2 [not reported]	--	[54]
LASA	1.2.2.1/2.4/3.1/4.3	83	OA (Netherlands)	Actigraph (<100 count/min)	0.78	+ 0.2 for 10 item -2.1 [-7.4 3.25] for 6 item	--	[33]
PAST	1.2.2.1/2.1/3.1/4.3	90	A (Australia)	activPAL (sitting/lying postures)	0.72	-1.0 [- 5.75 3.76]	t-test was inconclusive	[42]
MOST	1.2.2.1/2.2/3.2/4.5	48	OA (Australia)	Actigraph (<100 count/min)	0.67	-3.6 [-7.4 -0.2]	Guyatt Index 0.39 (0.47 for Actigraph)	[35]

272 A: adults; N: number of participants; OA: older adults; Ref: reference; UK: United Kingdom; USA: United States of America. For tool acronyms see table 1.

1  
2  
3 273 Criterion measure  
4  
5

6 274 None of the studies tested the accuracy of the tool against direct observation. Only two  
7  
8 275 studies.[16,42]used a postural sensor that actually measures sitting time objectively (activPAL), the  
9  
10 276 other twelve used an accelerometer built to measure low movement as a criterion measure  
11  
12 277 (ActiGraph, actiheart).  
13

14  
15  
16 278  
17

18 279 Statistical analysis  
19

20  
21 280 Accuracy and Limits of Agreement were usually derived from Bland and Altman plots. Sensitivity to  
22  
23 281 change was defined differently in the two articles which reported this measurement characteristic;  
24  
25 282 one used t-test statistics [42], one used the Guyatt Index [35].  
26  
27

28  
29 283  
30

31 284 Study Quality  
32

33  
34 285 Studies which scored highly for quality tended to be purposefully designed to test measurement  
35  
36 286 characteristics, rather than secondary analysis of data collected for another purpose. The most  
37  
38 287 common loss of quality was due to the use of accelerometers which assess low movement (e.g.  
39  
40 288 ActiGraph) as a criterion measure, as this does not measure the primary aspect of the definition of  
41  
42 289 SB (i.e. posture). Another issue which lowered quality was the manipulation of the criterion measure  
43  
44 290 without clear justification. For example, some studies manipulated the count threshold (used to  
45  
46 291 define SB) or included only SB bouts longer than a particular duration without justification or solid  
47  
48 292 rationale.  
49

50  
51  
52 293  
53

54  
55  
56 294 Tools and measurement characteristics  
57  
58  
59  
60

1  
2  
3 295 Table 3 summarises the results reported by these studies, arranged per measurement tool and  
4  
5 296 mapped against the relevant taxon. Very few of the existing tools to measure SB using self-report  
6  
7 297 have actually been investigated for these measurement characteristics. Accuracy has been reported  
8  
9 298 for six out of the 32 tools identified in the inventory (IPAQ-L 17d, IPAQ-S 17d, MOST, CHAMPS, LASA,  
10  
11 299 PAST). The most tested tool was the IPAQ in its long form, last seven days [16, 4-9] and short form,  
12  
13 300 last seven days.[5-3] CHAMPS was investigated in two studies [15, 53]. Information for other tools;  
14  
15 301 LASA [33], MOST [35], PAST [42], come from single studies. Reports of sensitivity to change are only  
16  
17 302 available for two tools; MOST [35] and PAST [42].  
18  
19  
20  
21 303  
22  
23  
24 304 Taxa tested  
25  
26  
27 305 The literature provides measurement characteristics information for six distinct full taxa:  
28  
29  
30 306 1.1.1/2.2/3.1/4.3 with five studies on IPAQ-L uw;  
31  
32  
33 307 1.1.1/2.2/3.1/4.1 with four studies on IPAQ-S uw  
34  
35  
36 308 1.2.2.1/2.3/3.2/4.5 with two studies on CHAMPS;  
37  
38  
39 309 1.2.2.1/2.4/3.1/4.3 with one study on LASA;  
40  
41  
42 310 1.2.2.1/2.1/3.1/4.3 with one study on PAST; and  
43  
44  
45 311 1.2.2.1/2.2/3.2/4.5 with one study on MOST.  
46  
47  
48 312 For the assessment type, there is information for direct measures via single item (1.1.1, nine studies)  
49  
50 313 and for composite sums of behaviours (1.2.2.1, five studies). However, there is no information for  
51  
52 314 direct proxy measures (1.1.2). For recall period, there is information on all four possible categories  
53  
54 315 (2.1 previous day, one study; 2.2 previous week, ten studies; 2.3 longer, three studies; and 2.4  
55  
56 316 unanchored, one study). The unanchored recall period (2.4), used by half of the tools in the  
57  
58  
59  
60

1  
2  
3 317 inventory, is particularly under-represented with only a single study in the validation review. For  
4  
5 318 temporal scale there is mostly information for assessment at day scale (3.1, twelve studies) and only  
6  
7 319 two studies for the temporal scale of a week (3.2). This is broadly representative of usage by tools in  
8  
9 320 the inventory. For assessment period there is information for weekdays only (4.1, four studies) or  
10  
11 321 both weekdays and weekend days (4.3, seven studies) and for tools with undefined assessment  
12  
13 322 periods (4.5, three studies). The not defined taxon (4.5) is under-represented by these validation  
14  
15 323 studies.

16  
17  
18  
19 324

20  
21  
22 325 Accuracy

23  
24  
25 326 Information for taxon 1.1.1/2.2/3.1/4.3 is not equivocal. The majority of studies reported a large  
26  
27 327 underestimation of total SB time ranging from 1.6 hours in adults [47] to 6 hours in older adults [46],  
28  
29 328 Others report that tools in this taxon overestimate total SB time by 2.2 hours in adults [48]. While  
30  
31 329 the direction of the error is equivocal it is clear that the systematic error on estimates of total SB  
32  
33 330 time using tools from this taxon is likely to be very large (several hours/day). The random error is  
34  
35 331 also likely to be very large as the Limits of Agreement reported were consistently very large.

36  
37  
38 332 Information for taxon 1.1.1/2.2/3.1/4.1 is a little more consistent for adults. Tools in this taxon seem  
39  
40 333 to underestimate total SB time by 1.5 to 3 hours in adults. However, in older adults this was less  
41  
42 334 clear with reports of underestimation by 4.4 hours [53] and overestimation by 3.5 hours [50]. In both  
43  
44 335 populations the error and Limits of Agreement were large, but not as large as for the previous taxon.

45  
46  
47 336

48  
49  
50 337 For taxon 1.2.2.1/2.3/3.2/4.5 there is consistent evidence for poor accuracy and large  
51  
52 338 underestimation of 5 to 7 hours/day in older adults. This implies that a longer recall period or the  
53  
54 339 temporal scale of a week might be less accurate. Other taxa under the 1.2.2.1 (sum of behaviours)  
55  
56 340 categories report smaller errors ranging from 0.2 hours overestimate [33] to a 3.6 hour

1  
2  
3 341 underestimation [35], both in older adults. One study [42], which used an objective measure of  
4  
5 342 sitting as criterion, reported a 1 hour underestimate and the smallest limit of agreement. This  
6  
7 343 suggests that tools in taxa with the attributes of 1.2.2.1 (sum of behaviours) and 3.1 (assessment at  
8  
9 344 day scale) are more likely to lead to accurate estimates of total SB time for a population.  
10

11  
12 345

13  
14  
15 346 Sensitivity to change  
16

17  
18 347 There is almost no information about sensitivity to change. The two studies that assessed sensitivity  
19  
20 348 to change [35, 42] provided little tangible information. The results were either inconclusive [42], or  
21  
22 349 reported the Guyatt index against a criterion measure which does not measure sitting [35]. While  
23  
24 350 the latter provided some indication that the tools' sensitivity to change was similar to that of an  
25  
26 351 objective measure of low movement it does not give a clear indication as to whether it is sensitive to  
27  
28 352 a change in total SB time. Neither of these studies reported the minimal detectable change [55], a  
29  
30 353 metric which provides an easily interpretable value of the capacity of a tool to detect a change.  
31  
32

33  
34 354  
35

## 36 37 355 **DISCUSSION** 38

39  
40 356 A taxonomy (TASST) for the systematic description and comparison of self-reported measures of SB  
41  
42 357 has been established. TASST provides a rigorous framework for informed choice, development and  
43  
44 358 evaluation of self-report tools. This framework has been used to review the measurement  
45  
46 359 characteristics of existing tools in order to identify the optimum tool for population surveillance. The  
47  
48 360 available evidence about measurement characteristics essential for population surveillance, namely  
49  
50 361 accuracy and responsiveness to change, was insufficient to ascertain which tool currently used in  
51  
52 362 practice is best. Accuracy was poor for all existing tools, with both under and over estimation of  
53  
54 363 total time spent in SB and large limits of agreement. In addition, there is a complete lack of evidence  
55  
56 364 about their sensitivity to change. Mapping available evidence onto the TASST framework has  
57  
58  
59  
60

1  
2  
3 365 enabled informed recommendations to be made about the promising features for a surveillance  
4  
5 366 tool, and identification of the aspects on which future research and development of SB surveillance  
6  
7 367 tools should focus.  
8  
9

10 368

11  
12  
13 369 The use of a coherent and robust taxonomy (TASST) to systematically evaluate and compare the  
14  
15 370 characteristics of measurement tools is the main strength of this study. However, in terms of  
16  
17 371 accuracy and sensitivity to change, the current published evidence does not cover the entire  
18  
19 372 taxonomy. Consequently, at present, only tentative recommendations can be provided. The  
20  
21 373 taxonomy can be used, however, to identify gaps in current research and provide focussed guidance  
22  
23 374 for future research and development. During the development of TASST, self-report tools which  
24  
25 375 aimed to measure SB in specific populations (e.g. children, those with arthritis) or specialised  
26  
27 376 contexts (e.g. workplace) were not considered. However, TASST is a generic framework, so tools  
28  
29 377 specific to these populations may already be fully described by the taxonomy. For example, a  
30  
31 378 question asking about time spent sitting at school which is specific to children, would be covered  
32  
33 379 under the sub-division of the day assessment period (taxon 4.4). Another consequence of the  
34  
35 380 exclusion criteria is that evidence on accuracy and sensitivity to change of tools specific to these  
36  
37 381 populations was not mapped on the taxonomy. Therefore, the conclusions drawn from the  
38  
39 382 measurement characteristics in this study are only valid for adults and older adults. In addition, this  
40  
41 383 study has the general limitations common to most systematic reviews, i.e. included articles were  
42  
43 384 restricted to those written in English, articles and tools published after the date of search were not  
44  
45 385 included, and any relevant articles not identified during the search will have been excluded.  
46  
47  
48  
49

50 386

51  
52  
53 387 The current study is the first to clearly define and focus on the measurement characteristics required  
54  
55 388 for population surveillance (accuracy and sensitivity to change). There is only one other systematic  
56  
57  
58  
59  
60

1  
2  
3 389 review reporting on the measurement characteristics of self-report tools to measure SB [13], which  
4  
5 390 concentrated on validity (assessed through rank correlation) and reliability, which are the  
6  
7 391 measurement characteristics relevant to establishing association between SB and health. In  
8  
9 392 agreement with the previous review, we found that the major flaw of most validation studies was  
10  
11 393 the use of an inadequate criterion measure. The choice of criterion measure depends on the  
12  
13 394 purpose of the tool. While direct observation should be considered the gold standard, if the purpose  
14  
15 395 is to assess total sedentary time, then accurate postural sensors should be adequate (e.g. activPAL).  
16  
17 396 Instead, many studies used an accelerometer which measures low levels of movement at the hip  
18  
19 397 (e.g. ActiGraph) as a criterion measure, but such tools do not measure SB directly and can misclassify  
20  
21 398 standing as sitting [12].  
22  
23  
24  
25  
26  
27

28 400 Despite the incomplete nature of the evidence, TASST enables the identification of desirable  
29  
30 401 characteristics of self-report tools to measure SB when used for population surveillance. Firstly,  
31  
32 402 tools assessing total SB time as a sum of behaviours (taxon 1.2.2.1) provided better accuracy than  
33  
34 403 single item (taxon 1.1) tools. However, this will be dependent on the behaviours or domains  
35  
36 404 included within the sum, and whether they are exhaustive, consistent and mutually exclusive. Tools  
37  
38 405 with a non-exhaustive sum will underestimate total time, for example, the Longitudinal Aging Study  
39  
40 406 Amsterdam (LASA), found that a six item sum provided a better correlation with SB across the  
41  
42 407 sample, but that a ten item sum was more accurate [33]. Conversely, tools which contain  
43  
44 408 behaviours which might occur concurrently (such as watching TV and using a tablet computer) may  
45  
46 409 lead to an over-estimate in total SB time [56]. Secondly, tools using a previous day recall period  
47  
48 410 (taxon 2.1) provide better accuracy than those with longer recall periods (taxa 2.2 and 2.3). This  
49  
50 411 corroborates recent research on the validity of computerised survey systems which assess SB using a  
51  
52 412 past-day recall period [17, 57]. However, although tools using previous day recall may more  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 413 accurate, it is likely that their sensitivity to change will be less good due to the higher underlying  
4  
5 414 variability in daily SB [58].  
6  
7  
8 415  
9  
10  
11 416 Most tools currently used for population surveillance of SB systematically underestimate the amount  
12  
13 417 of SB by two to four hours per day. Yet, self-report tools are still the most practical and economical  
14  
15 418 means of population surveillance. Therefore, policy makers and clinicians should be aware that  
16  
17 419 reports of population SB time are likely to be grossly underestimated, and should be cognisant of  
18  
19 420 this fact when making decisions on implementing, developing and evaluating policy and public  
20  
21 421 health interventions. In addition, policy makers and clinicians should be cautious in interpreting any  
22  
23 422 reported difference in population SB time as a real change. The dearth of information about  
24  
25 423 sensitivity to change of these tools means that we do not know the magnitude of change required to  
26  
27 424 be certain that a change is real and not background variation. Moving forward, development of  
28  
29 425 national and international surveillance systems should not be undertaken assuming that a tool is  
30  
31 426 adequate because it has been used previously. Instead, investment should be made in research to  
32  
33 427 evaluate the sensitivity to change and accuracy of tools to measure SB, paying attention to the  
34  
35 428 potential trade-off between these two measurement characteristics. Such research should be  
36  
37 429 carefully planned, to ensure that meaningful comparisons are investigated. The TASST taxonomy  
38  
39 430 should be used as a useful framework to facilitate such a systematic approach.  
40  
41  
42  
43  
44 431  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 432 Acknowledgements  
4  
5

6 433 The named authors present the study on behalf of the Seniors USP Team, which comprises: Dawn A  
7  
8 434 Skelton (PI), Sebastien Chastin, Simon Cox, Elaine Coulter, Iva Cukic, Philippa Dall, Ian Deary, Geoff  
9  
10 435 Der, Manon Dontje, Claire Fitzsimons, Catherine Gale, Jason Gill, Malcolm Granat, Cindy Gray,  
11  
12 436 Carolyn Greig, Elaine Hindle, Karen Laird, Gillian Mead, Nanette Mutrie, Victoria Palmer, Ratko  
13  
14 437 Radakovic, Naveed Sattar, Richard Shaw, John Starr, Sally Stewart, Sally Wyke.  
15  
16

17  
18 438  
19

20 439 Competing Interests:  
21  
22

23 440 All authors had financial support from the Medical Research Council for the submitted work; no  
24  
25 441 financial relationships with any organisations that might have an interest in the submitted work in  
26  
27 442 the previous three years; no other relationships or activities that could appear to have influenced  
28  
29 443 the submitted work.  
30  
31

32  
33 444  
34  
35

36 445 Funding:  
37  
38

39 446 The study was performed in preparation for the Seniors USP: Understanding Sedentary Patterns  
40  
41 447 project, which was supported by the Medical Research Council [MR/K025023/1]  
42  
43

44 448  
45  
46

47 449 Authors Contributions:  
48  
49

50 450 SC and DS conceived and supervised the study. PD and SC designed the study. PD, EC, CF and SC  
51  
52 451 collected the data, and performed the systematic review. PD, CF and SC analysed the data for the  
53  
54 452 taxonomy. PD and SC analysed the data from the systematic review. All authors interpreted the  
55  
56 453 data and critically revised the manuscript for important intellectual content.  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

454

455 Data Sharing: Data can be obtained from the corresponding author on request.

456

457 Figure Legends

458

459 Figure 1: Taxonomy of Self-reported Sedentary behaviour Tools (TASST)

460

461 Figure 2: PRISMA diagram of the validation systematic review

462

For peer review only

1  
2  
3 463 References  
4  
5  
6 464  
7  
8  
9 465 1. Kohl HW 3rd, Craig CL, Lambert EV, et al. The pandemic of physical inactivity: global action for  
10  
11 466 public health. *Lancet* 2012;380:294–305.  
12  
13  
14 467 2. Chastin S, Schwartz U, Skelton D. Development of a consensus taxonomy of sedentary behaviors  
15  
16 468 (SIT): report of Delphi round 1. *PLoS One* 2013;8:e82313.  
17  
18  
19 469 3. Sedentary Behaviour Research Network. Standardized use of the terms “sedentary” and  
20  
21 470 “sedentary behaviours”. *Appl Physiol Nutr Metabol* 2012;37:540-2.  
22  
23  
24 471 4. Healy GN, Matthews CE, Dunstan DW, Winkler EAH, Owen N. Sedentary time and cardio-  
25  
26 472 metabolic biomarkers in US adults: NHANES 2003-06. *Eur Heart J* 2011;32:590-7.  
27  
28  
29 473 5. Harvey JA, Chastin SFM, Skelton DA. How sedentary are older people? A systematic review of the  
30  
31 474 amount of sedentary behavior. *J Aging Phys Act* 2015;23:471-87.  
32  
33  
34 475 6. Proper KI, Singh AS, van Mechelen W, Chinapaw MJM. Sedentary behaviors and health outcomes  
35  
36 476 among adults: a systematic review of prospective studies. *Am J Prev Med* 2011;40:174-82.  
37  
38  
39 477 7. Thorp AA, Owen N, Neuhaus M, Dunstan DW. Sedentary behaviors and subsequent health  
40  
41 478 outcomes in adults a systematic review of longitudinal studies, 1996-2011. *Am J Prev Med*  
42  
43 479 2011;41:207-15.  
44  
45  
46 480 8. de Rezende LFM, Rey-López JP, Matsudo VKR, do Carmo Luiz O. Sedentary behavior and health  
47  
48 481 outcomes among older adults: a systematic review. *BMC Public Health* 2014;14:333.  
49  
50  
51 482 9. Dogra S, Stathokostas L. Sedentary behavior and physical activity are independent predictors of  
52  
53 483 successful aging in middle-aged and older adults. *J Aging Res* 2012;2012:190654.  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 484 10. Department of Health. Start Active, Stay Active: A report on physical activity from the four home  
4  
5 485 countries' Chief Medical Officers. 2011.  
6  
7  
8 486 11. Guyatt G, Walter S, Norman G. Measuring change over time. assessing the usefulness of  
9  
10 487 evaluative instruments. *J Chron Dis* 1987;40:171-8.  
11  
12  
13 488 12. Kozey-Keadle S, Libertine A, Lyden K, Staudenmayer J, Freedson P. Validation of wearable  
14  
15 489 monitors for assessing sedentary behavior. *Med Sci Sports Exerc* 2011;43:1561-7.  
16  
17  
18 490 13. Healy GN, Clark BK, Winkler EAH, Gardiner PA, Brown WJ, Matthews CE. Measurement of adults'  
19  
20 491 sedentary time in population-based studies. *Am J Prev Med* 2011;41:216-27.  
21  
22  
23 492 14. Bauman A, Ainsworth BE, Sallis JF, et al. The descriptive epidemiology of sitting: a 20 country  
24  
25 493 comparison using the International Physical Activity Questionnaire (IPAQ). *Am J Prev Med*  
26  
27 494 2011;41:228-35.  
28  
29  
30 495 15. Hekler EB, Buman MP, Haskell WL, et al. Reliability and validity of CHAMPS self-reported  
31  
32 496 sedentary-vigorous intensity physical activity in older adults. *J Phys Act Health* 2012;9:225-36.  
33  
34  
35 497 16. Chastin SFM, Culhane B, Dall PM. Questionnaire sitting time (IPAQ) compared to objective  
36  
37 498 measurement via inclinometry (activPAL). *Physiol Meas* 2014;35:2319-28.  
38  
39  
40 499 17. Kozey-Keadle S, Lyden K, Hickey A, et al. Validation of a previous day recall for measuring the  
41  
42 500 location and purpose of active and sedentary behaviours compared to direct observation. *Int J*  
43  
44 501 *Behav Nutr Phys Act* 2014;11:12.  
45  
46  
47 502 18. Kmet LM, Lee RC, Cook LS. Standard quality assessment criteria for evaluating primary research  
48  
49 503 papers from a variety of fields. *Edmonton: Alberta Heritage Foundation for Medical Research*  
50  
51 504 2004;1-22.  
52  
53  
54 505 19. Banks E, Jorm L, Rogers K, Clements M, Bauman A. Screen-time, obesity, ageing and disability:  
55  
56 506 findings from 91 266 participants in the 45 and Up Study. *Public Health Nutr* 2011;14:34-43.  
57  
58  
59  
60

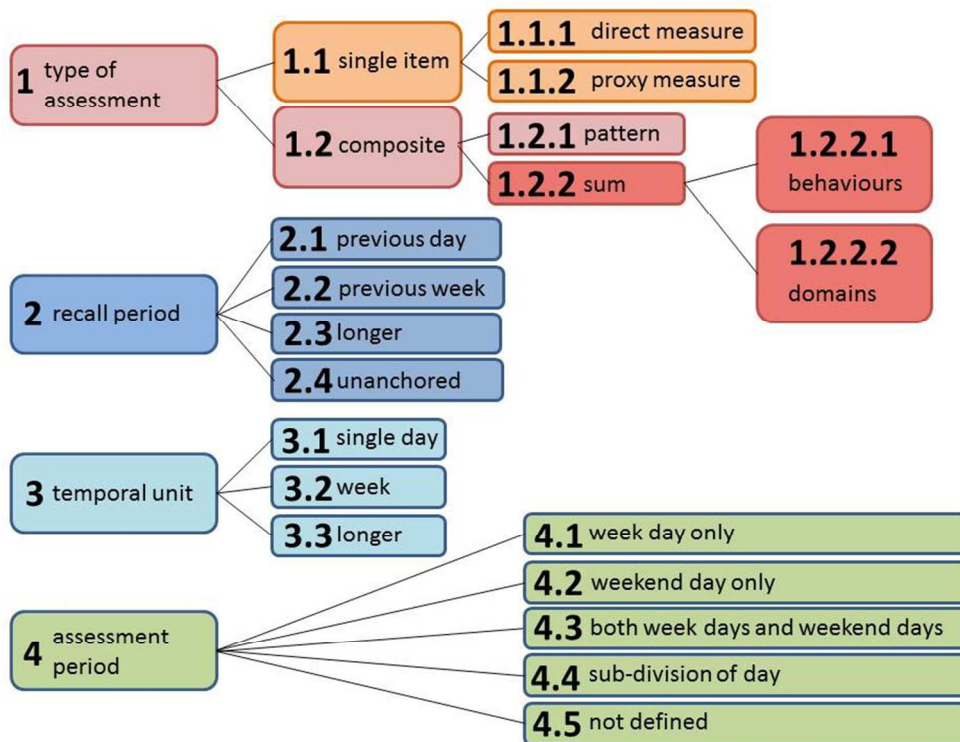
- 1  
2  
3 507 20 Patel AV, Bernstein L, Deka A, et al. Leisure time spent sitting in relation to total mortality in a  
4  
5 508 prospective cohort of US adults. *Am J Epidemiol* 2010;172:419-429.  
6  
7  
8 509 21. Salmon J, Owen N, Crawford D, Bauman A, Sallis JF. Physical activity and sedentary behavior: a  
9  
10 510 population-based study of barriers, enjoyment, and preference. *Health Psychol* 2003;22:178-88.  
11  
12  
13 511 22. Gardiner PA, Healy GN, Eakin EG, et al. Associations between television viewing time and overall  
14  
15 512 sitting time with the metabolic syndrome in older men and women: The Australian Diabetes Obesity  
16  
17 513 and Lifestyle study. *JAGS* 2011;59:788-96.  
18  
19  
20 514 23. Shields M, Tremblay MS. Sedentary behaviour and obesity. *Health Rep.* 2008;19:19-30.  
21  
22  
23 515 24. Katzmarzyk PT, Church TS, Craig Cl, Bouchard C. Sitting time and mortality from all causes,  
24  
25 516 cardiovascular disease, and cancer. *Med Sci Sports Exerc* 2009;41:998-1005.  
26  
27  
28 517 25. Hamer M, Stamatakis E. Screen-based sedentary behaviour, physical activity, and muscle  
29  
30 518 strength in the English Longitudinal Study of Ageing. *PLoS One* 2013;8:e66222.  
31  
32  
33 519 26. Wareham NJ, Jakes RW, Rennie KL, Mitchell J, Hennings S, Day NE. Validity and repeatability of  
34  
35 520 the EPIC-Norfolk physical activity questionnaire. *Int J Epidemiol* 2002;31:186-74.  
36  
37  
38 521 27. Parsons TJ, Thomas C, Power C. Estimated activity patterns in British 45 year olds: cross-  
39  
40 522 sectional findings from the 1958 British birth cohort. *Eur J Clin Nutr* 2009;63:978-85.  
41  
42  
43 523 28. Armstrong T, Bull F. Development of the World Health Organization Global Physical Activity  
44  
45 524 Questionnaire (GPAQ). *J Public Health* 2006;14:66-70.  
46  
47  
48 525 29. Stamatakis E, Davis M, Stathi A, Hamer M. Associations between multiple indicators of  
49  
50 526 objectively-measured and self-reported sedentary behaviour and cardiometabolic risk in older  
51  
52 527 adults. *Prev Med* 2012;54:82-7.  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 528 30. Chau JY, Grunseit A, Midthjell K, et al. Sedentary behaviour and risk of mortality from all-causes  
4  
5 529 and cardiometabolic diseases in adults: evidence from the HUNT3 population cohort. *Br J Sports*  
6  
7 530 *Med* 2015;49:737-42.  
8  
9  
10 531 31. Craig CL, Marshall AL, Sjöström M, et al. International Physical Activity Questionnaire: 12-country  
11  
12 532 reliability and validity. *Med Sci Sports Exerc* 2003;35:1381-95.  
13  
14  
15 533 32. Proper KI, Cerin E, Brown WJ, Owen N. Sitting time and socio-economic differences in overweight  
16  
17 534 and obesity. *Int J Obes* 2007;31:169-76.  
18  
19  
20 535 33. Visser M, Koster A. Development of a questionnaire to assess sedentary time in older persons – a  
21  
22 536 comparative study using accelerometry. *BMC Geriatr* 2013;13:80.  
23  
24  
25 537 34. Burazeri G, Goda A, Kark JD. Television viewing, leisure-time exercise and acute coronary  
26  
27 538 syndrome in transitional Albania. *Prev Med* 2008;47:112-5.  
28  
29  
30 539 35. Gardiner PA, Clark BK, Healy GN, Eakin EG, Winkler EAH, Owen N. Measuring older adults'  
31  
32 540 sedentary time: reliability, validity, and responsiveness. *Med Sci Sports Exerc* 2011;43: 2127-33.  
33  
34  
35 541 36. Ford ES. Combined television viewing and computer use and mortality from all-causes and  
36  
37 542 diseases of the circulatory system among adults in the United States. *BMC Public Health* 2012;12:70  
38  
39  
40 543 37. Evenson KR, McGinn AP. Test-retest reliability of adult surveillance measures for physical activity  
41  
42 544 and inactivity. *Am J Prev Med* 2005;28:470-8.  
43  
44  
45 545 38. Wolf AM, Hunter DJ, Colditz GA, et al. Reproducibility and validity of a self-administered physical  
46  
47 546 activity questionnaire. *Int J Epidemiol* 1994;23:991-99.  
48  
49  
50 547 39. Matthews CE, George SM, Moore SC, et al. Amount of time spent in sedentary behaviors and  
51  
52 548 cause-specific mortality in US adults. *Am J Clin Nutr* 2012;ajcn-019620.  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 549 40. Salmon J, Bauman A, Crawford D, Timperio A, Owen N. The association between television  
4  
5 550 viewing and overweight among Australian adults participating in varying levels of leisure-time  
6  
7 551 physical activity. *Int J Obes Relat Metab Disord* 2000;24:600-6.  
8  
9  
10 552 41. Washburn RA, McAuley E, Katula J, Mihalko SL, Boileau RA. The Physical Activity Scale for the  
11  
12 553 Elderly (PASE): evidence for validity. *J Clin Epidemiol* 1999;52:643-51.  
13  
14  
15 554 42. Clark BK, Winkler E, Healy GN, et al. Adults' past-day recall of sedentary time: reliability, validity  
16  
17 555 and responsiveness. *Med Sci Sports Exerc* 2013;45:1198-207.  
18  
19  
20 556 43. Balboa-Castillo T, León-Muñoz LM, Graciani A, Rodríguez-Artalejo F, Guallar-Castillón P.  
21  
22 557 Longitudinal association of physical activity and sedentary behavior during leisure time with health-  
23  
24 558 related quality of life in community-dwelling older adults. *Health Qual Life Outcomes* 2011;9:41-47.  
25  
26  
27 559 44. Rosenberg DE, Norman GK, Wagner N, Patrick K, Calfas KJ, Sallis JF. Reliability and validity of the  
28  
29 560 Sedentary Behaviour Questionnaire (SBQ) for adults. *J Phys Act Health* 2010;7:697-705.  
30  
31  
32 561 45. Stamatakis E, Hamer M, Dunstan DW. Screen-based entertainment time, all-cause mortality, and  
33  
34 562 cardiovascular events. *J Am Coll Cardiol* 2011;57:292-9.  
35  
36  
37 563 46. Dahl-Pederson KI, Hansen AW, Bjerregaard P, Jørgensen ME, Brage S. Validity of the  
38  
39 564 international physical activity questionnaire in the arctic. *Med Sci Sports Exerc* 2013;45:728-36.  
40  
41  
42 565 47. van Dyck D, Cardon G, Deforche B, de Bourdeaudhuij I. IPAQ interview version: convergent  
43  
44 566 validity with accelerometers and comparison of physical activity and sedentary time levels with the  
45  
46 567 self-administered version. *J Sports Med Phys Fitness* 2014;55:776-86.  
47  
48  
49 568 48. Hagstromer M, Ainsworth BE, Sjostrom M. Comparison of a subjective and an objective measure  
50  
51 569 of physical activity in a population sample. *J Phys Act Health* 2010;7:541-50.  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 570 49. Celis-Morales CA, Perez-Bravo F, Ibanez L, Salas C, Bailey MES, Gill JMR. Objective vs. self-  
4  
5 571 reported physical activity and sedentary time: effects of measurement method on relationships with  
6  
7 572 risk biomarkers. *PLoS ONE* 2012;7:e36345.  
8  
9  
10 573 50. Dyrstad SM, Hansen BH, Holme IM, Anderssen SA. Comparison of self-reported versus  
11  
12 574 accelerometer-measured physical activity. *Med Sci Sports Exerc* 2014;46:99-106.  
13  
14  
15 575 51. Oyeyemi AL, Umar M, Oguche F, Aliyu SU, Oyeyemi AY. Accelerometer-determined physical  
16  
17 576 activity and its comparison with the International Physical Activity Questionnaire in a sample of  
18  
19 577 Nigerian adults. *PLoS ONE* 2014;9:e87233.  
20  
21  
22 578 52. Hurtig-Wennlöf A, Hagströmer M, Olsson L. The International Physical Activity Questionnaire  
23  
24 579 modified for the elderly: aspects of validity and feasibility. *Public Health Nutr* 2010;13:1847-54.  
25  
26  
27 580 53. Grimm EK, Swartz AM, Hart T, Miller NE, Strath SJ. Comparison of the IPAQ-short form and  
28  
29 581 accelerometry predictions of physical activity in older adults. *J Aging Phys Act* 2012;20:64-79.  
30  
31  
32 582 54. Gennuso KP, Matthews CE, Colbert LH. Reliability and validity of two self-report measures to  
33  
34 583 assess sedentary behavior in older adults. *J Phys Act Health* 2015;12:727-32.  
35  
36  
37 584 55. Beaton DE, Bombardier C, Katz JN, Wright JG. A taxonomy for responsiveness. *J Clin Epidemiol*  
38  
39 585 2001;54:1204-71.  
40  
41  
42 586 56. Busschaert C, de Bourdeaudhuij I, van Holle V, Chastin SFM, Cardon G, de Cocker K. Reliability  
43  
44 587 and validity of three questionnaires measuring context-specific sedentary behaviour and associated  
45  
46 588 correlates in adolescents, adults, and older adults. *Int J Behav Nutr Phys Act* 2015;12:117.  
47  
48  
49 589 57. Gomersall SR, Pavey T, Clark BK, Jasman A, Brown WJ. Validity of a self-report recall tool for  
50  
51 590 estimating sedentary behaviour in adults. *J Phys Act Health* 2015;12:1485-91.  
52  
53  
54 591 58. Hart TL, Swartz AM, Cashin SE, Strath SJ. How many days of monitoring predict physical activity  
55  
56 592 and sedentary behaviour in older adults? *Int J Behav Nutr Phys Act* 2011;8:62.  
57  
58  
59  
60





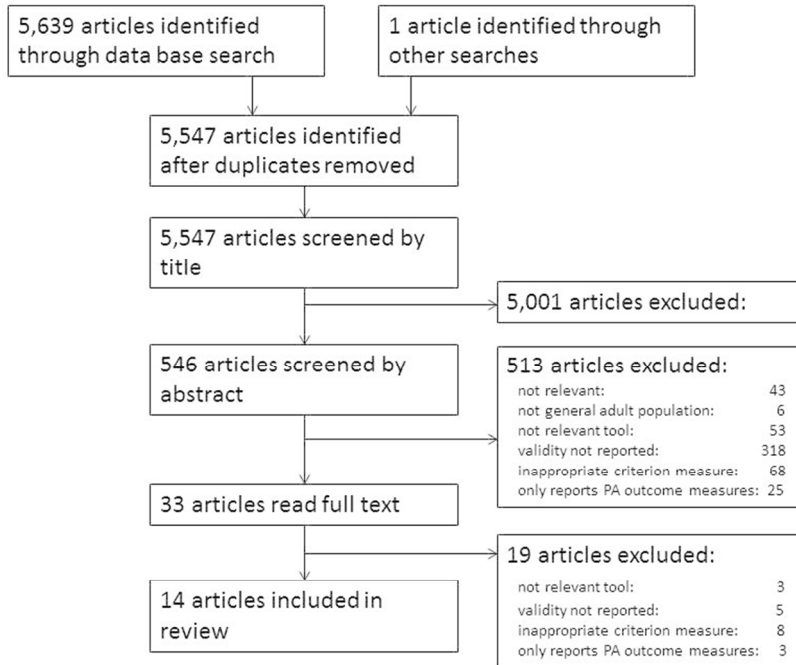
Taxonomy of Self-reported Sedentary behaviour Tools (TASST)

Figure 1

254x190mm (96 x 96 DPI)

ew only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



PRISMA diagram of the validation systematic review  
Figure 2  
254x190mm (96 x 96 DPI)

ew only



# PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
<b>TITLE</b>			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
<b>ABSTRACT</b>			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
<b>INTRODUCTION</b>			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	3
<b>METHODS</b>			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	4
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	4 & 5
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	4 & 5
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	5
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	5
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	5
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	5
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	Not reporting on trials, so not done



# PRISMA 2009 Checklist

Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	5
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., $I^2$ ) for each meta-analysis.	Not done

Page 1 of 2

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	9
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	Not done
<b>RESULTS</b>			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	8, figure 2
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Table 3
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	8
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Table 3, pages 9 & 10 (not a review of interventions)
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	Not enough data for meta-analysis
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	9
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	Not done
<b>DISCUSSION</b>			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	12
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	11 & 12



# PRISMA 2009 Checklist

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	12
<b>FUNDING</b>			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	14

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: [www.prisma-statement.org](http://www.prisma-statement.org).

Page 2 of 2

# BMJ Open

## Taxonomy of Self-reported Sedentary behaviour Tools (TASST): a framework for development, comparison and evaluation

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2016-013844.R1
Article Type:	Research
Date Submitted by the Author:	30-Nov-2016
Complete List of Authors:	Dall, Philippa; Glasgow Caledonian University School of Health and Life Sciences, Department of Psychology, Social Work & Allied Health Sciences Coulter, Elaine; University of Glasgow Fitzsimons, Claire F.; Univ Edinburgh, ISPEHS Skelton, Dawn; Glasgow Caledonian University, School of Health & Life Sciences Chastin, Sebastian; Glasgow Caledonian University
<b>Primary Subject Heading</b>:	Public health
Secondary Subject Heading:	Epidemiology, Research methods, Sports and exercise medicine
Keywords:	sedentary behaviour, sitting, population surveillance, measurement, validation

SCHOLARONE™  
Manuscripts

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1 **Title:** TAxonomy of Self-reported Sedentary behaviour Tools (TASST): a framework for development,  
2 comparison and evaluation

3

4 **Authors:**

5 Dall PM<sup>1</sup>, Coulter EH<sup>2</sup>, Fitzsimons CF<sup>3</sup>, Skelton DA<sup>1</sup>, Chastin SFM<sup>1</sup>, on behalf of the Seniors USP Team.

6

7 <sup>1</sup> Institute for Applied Health Research, Glasgow Caledonian University, Cowcaddens Road, Glasgow,  
8 G4 0BA

9 <sup>2</sup> Department of Nursing and Healthcare, Glasgow University, 59 Oakfield Avenue, Glasgow, G12 8LL

10 <sup>3</sup> Sport, Physical Education and Health Sciences Institute, Edinburgh University, St Leonard's Land,  
11 Holyrood Road, Edinburgh, EH8 8AQ

12

13 corresponding author:

14 Dr Philippa Dall

15 School of Health & Life Sciences

16 Glasgow Caledonian University

17 Cowcaddens Road

18 Glasgow

19 G4 0BA

20 philippa.dall@gcu.ac.uk

21

22 Word count: 5,035

23

1  
2  
3 24 **ABSTRACT**  
4  
5

6 25 **Objective:** Sedentary behaviour (SB) has distinct deleterious health outcomes, yet there is no  
7  
8 26 consensus on best practice for measurement. This study aimed to identify the optimal self-report  
9  
10 27 tool for population surveillance of SB, using a systematic framework.  
11

12  
13 28 **Design:** A framework, Taxonomy of Self-report SB Tools (TASST), consisting of four domains (type of  
14  
15 29 assessment, recall period, temporal unit, and assessment period), was developed based on a  
16  
17 30 systematic inventory of existing tools. The inventory was achieved through a systematic review of  
18  
19 31 studies reporting SB and tracing back to the original description. A systematic review of the accuracy  
20  
21 32 and sensitivity to change of these tools was then mapped against TASST domains.  
22  
23

24  
25 33 **Data Sources:** Systematic searches were conducted via EBSCO, reference lists and expert opinion.  
26  
27

28 34 **Eligibility Criteria for selecting studies:** The inventory included tools measuring SB in adults that  
29  
30 35 could be self-completed at one sitting, and excluded tools measuring SB in specific populations or  
31  
32 36 contexts. The systematic review included studies reporting on the accuracy against an objective  
33  
34 37 measure of SB and/or sensitivity to change of a tool in the inventory.  
35  
36

37 38 **Results:** The systematic review initially identified 32 distinct tools (141 questions), which were used  
38  
39 39 to develop the TASST framework. Twenty-two studies evaluated accuracy and/or sensitivity to  
40  
41 40 change representing only 8 taxa. Assessing SB as a sum of behaviours and using a previous day recall  
42  
43 41 were the most promising features of existing tools. Accuracy was poor for all existing tools, with  
44  
45 42 both under and over estimation of SB. There was a lack of evidence about sensitivity to change.  
46  
47

48  
49 43 **Conclusions:** Despite the limited evidence, mapping existing SB tools onto the TASST framework has  
50  
51 44 enabled informed recommendations to be made about the most promising features for a  
52  
53 45 surveillance tool, identified aspects on which future research and development of SB surveillance  
54  
55 46 tools should focus.  
56  
57  
58  
59  
60



1  
2  
3 47 **Systematic Review Registration**  
4  
5

6 48 PROSPERO/CRD42014009851  
7  
8

9 49  
10

11  
12 50 **KEY WORDS:**  
13

14  
15 51 sedentary behaviour; sitting; population surveillance; measurement; validation  
16  
17

18 52  
19

20  
21 53 **STRENGTHS AND LIMITATIONS OF THIS STUDY:**  
22

- 23  
24 54 • A systematic approach was taken towards classifying self-reported measures of sedentary  
25  
26 55 behaviour, allowing a structured approach to measurement in the future  
27  
28 56 • An example of use of the framework is presented, mapping accuracy and sensitivity to  
29  
30 57 change of self-report sedentary behaviour measures on to the framework  
31  
32 58 • Although designed to be generic, the TASST framework was developed excluding tools  
33  
34 59 measuring sedentary behaviour in specialised populations and contexts, e.g. children or the  
35  
36 60 workplace, and the framework may therefore not cover some aspects of these tools  
37  
38 61 • There is the potential for a language bias, as full-text articles not in English were not  
39  
40 62 included in the systematic reviews.  
41  
42  
43  
44 63  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 64 **BACKGROUND**  
4  
5

6 65 Physical inactivity is currently at pandemic levels [1] and is a global public health concern. Sedentary  
7  
8 66 behaviour (SB), an umbrella term for all waking time spent in non-exercising sitting or reclining  
9  
10 67 postures [2, 3] such as sitting during work, motorised transport or watching TV, is the largest  
11  
12 68 contributor to inactivity [4,5]. Higher levels of SB have been associated with poor physical and  
13  
14 69 mental health, increased risk of chronic disease and less successful ageing [6-9]. Consequently,  
15  
16 70 several countries, including the UK, have issued recommendations to reduce SB at all ages as part of  
17  
18 71 their national physical activity guidelines [10]. Population surveillance is urgently needed to monitor  
19  
20 72 the impact of such policy, track changes in SB over time, and to evaluate public health interventions  
21  
22 73 targeting SB. In order to provide effective surveillance upon which to base future policy decisions,  
23  
24 74 such surveillance tools should be accurate (provide a true measure of the actual amount of SB in a  
25  
26 75 population) and sensitive to change (provide the true difference in SB between two measurement  
27  
28 76 time points) [11].  
29  
30  
31  
32  
33  
34  
35

36 78 Objective body worn sensors, that measure posture, demonstrate good accuracy for measuring total  
37  
38 79 duration of SB against the gold standard of direct observation [12], but they are expensive and  
39  
40 80 challenging to use for population surveillance. Self-report tools provide a pragmatic choice for  
41  
42 81 population surveillance and have the potential to provide context rich information, useful for  
43  
44 82 intervention development [13]. To date, surveys assessing SB have predominantly used self-report  
45  
46 83 tools [14], which are generally adapted from tools not specifically designed to measure that  
47  
48 84 behaviour (e.g. tools designed to measure physical activity) [15], and which have not been evaluated  
49  
50 85 for population surveillance purposes [14]. No framework currently exists with which to describe and  
51  
52 86 compare SB self-report tools, meaning there is currently no way of systematically selecting an  
53  
54 87 appropriate tool. A previous systematic review of the measurement characteristics of self-report  
55  
56 88 tools measuring SB, reported acceptable to good reliability but low to moderate correlation with a  
57  
58  
59  
60

1  
2  
3 89 (non-gold standard) criterion measure [13]. This suggests that self-report measures of SB are  
4  
5 90 acceptable tools to establish epidemiological evidence of an association between SB and health [13].  
6  
7 91 However, it is possible that the scale of the problem may be vastly underestimated, as differences of  
8  
9 92 2-4 hours per day (approximately 20% of SB) have been reported between self-report and objective  
10  
11 93 tools [16].  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40

94  
95 The primary aim of this study was to identify, in a systematic manner, the optimal self-report tool to  
96 measure SB for use in population surveillance. Although self-report SB tools can and will be used in  
97 other areas of research, this study focussed on population surveillance as an area that is crucial to  
98 the development of public health policy. To fulfil the primary aim, a framework was created to  
99 describe the features of self-report tools measuring SB, the TAXonomy of Self-report Sedentary  
100 behaviour Tools (TASST). A systematic inventory of existing self-report tools to measure SB was  
101 mapped onto TASST, and the measurement characteristics of these tools, focussing on accuracy and  
102 sensitivity to change, were evaluated, with explicit reference to the domains of the taxonomy  
103 framework.  
104

## 105 **METHODS**

106 The study protocol (PROSPERO CRD42014009851), was conducted in three phases. In phase 1 an  
107 exhaustive inventory of self-reported tools to measure SB in adults and older adults was established  
108 using a structured search protocol. Phase 2 was the development of a taxonomy based on content  
109 analysis of the items and questions in the tools. In phase 3, a systematic literature review of the  
110 measurement characteristics of the tools in the inventory was conducted and mapped onto the  
111 taxonomy.  
112

1  
2  
3 113 **Phase 1: Systematic inventory of self-report tools**  
4  
5

6 114 The aim of the systematic inventory was to compile an exhaustive list of self-report tools which  
7  
8 115 could be used to measure SB in adults ( $\geq 18$  years) and older adults ( $\geq 60$  years). Since the aim was to  
9  
10 116 identify tools and not to identify articles, this stage does not have the same methodology as a  
11  
12 117 systematic literature review. A literature search was conducted in October 2013 (updated  
13  
14 118 November 2016), for articles reporting SB as an outcome measure. From this review, a list of self-  
15  
16 119 report tools which measured SB was compiled. References lists were reviewed and experts  
17  
18 120 consulted to identify any additional tools to include in the inventory. The inventory then was  
19  
20 121 consolidated to amalgamate tools referred to by different names, and to trace back to the original  
21  
22 122 version. Articles which added significant new questions to tools were included as a separate tool.  
23  
24 123 We defined significant new questions to be at least one question which added or changed the type  
25  
26 124 of sedentary behaviour or the time period considered by the tool. Changes in phrasing of the  
27  
28 125 question were not considered sufficient to be considered as a separate tool. Tools used in a single  
29  
30 126 study and those without names/acronyms were included as separate tools.  
31  
32  
33  
34  
35  
36  
37

38 127

39 128 To be included in the inventory, tools had to: be suitable for use for large scale population studies of  
40  
41 129 adults or older adults, including being suitable for self-completion by the respondent at a single  
42  
43 130 point in time (a pragmatic requirement to minimise participant burden); and measure SB or a proxy  
44  
45 131 measure of SB (e.g. TV viewing). Tools were excluded from the inventory: if they were designed  
46  
47 132 specifically to assess SB in children or other specialised populations (e.g. medical conditions); if they  
48  
49 133 were designed specifically to assess SB in a specialised context, (e.g. workplace or care settings); if  
50  
51 134 continuous reporting over extended periods of time was required (e.g. diaries or time-use surveys);  
52  
53 135 or if significant interviewer interactions were required. Self-report tools that could be administered  
54  
55 136 by telephone or interview were not automatically excluded, however tools such as the PDR (Previous  
56  
57  
58  
59  
60

1  
2  
3 137 Day Recall) [17], in which the interviewer works through lists of several hundred items, were  
4  
5 138 excluded.

6  
7  
8 139

9  
10  
11 140 **Phase 2: Development of a taxonomy**

12  
13  
14 141 Only tools identified in the initial search were used to develop the taxonomy. The original text was  
15  
16 142 extracted for each question relating to SB in each of the self-report tools identified in the inventory.  
17  
18 143 Content analysis was conducted on the text to extract all of the attributes in the questions that were  
19  
20 144 used to describe and constrain what aspect of SB was measured by that question. For example, in  
21  
22 145 the question “During the last 7 days, how much time did you usually spend sitting on a week day?”,  
23  
24 146 attributes extracted relating to the measurement of SB would be “during the last 7 days”, “time  
25  
26 147 spent sitting” and “on a week day”. Attributes were then grouped into mutually exclusive domains  
27  
28 148 covering similar aspects of measurement, and categories within those domains were defined  
29  
30 149 iteratively. A new category was created each time a tool did not fit within an existing category. The  
31  
32 150 full taxonomy was then assembled and streamlined by merging categories with overlapping  
33  
34 151 meaning. Finally, consideration was given to potential future developments of self-report tools to  
35  
36 152 measure SB, such as the growing interest in the pattern of accumulation of sedentary behaviour, by  
37  
38 153 adding any categories to the taxonomy considered useful in the future. The resulting taxonomy was  
39  
40 154 then tested by ensuring that all tools could be classified similarly by two independent researchers  
41  
42 155 and that the taxonomy fully defined the tool.  
43  
44  
45  
46  
47  
48  
49

50

51 157 **Phase 3: Systematic review of measurement characteristics**

52  
53 158 Finally, a systematic literature search in relevant health databases was conducted in December 2014  
54  
55 159 (updated November 2016) via EBSCO host. The search combined the name of the tool including  
56  
57 160 variants and acronyms (except where the acronym was also a common word, e.g. PAST, MOST), with  
58  
59  
60

1  
2  
3 161 search terms relating to measurement characteristics (valid\* /reliab\* /repons\* /sensitiv\* /calibrat\*  
4  
5 162 /accura\* /agreement /psychometric\* /clinimetric\* /"measurement characteristics" /Reliability and  
6  
7 163 Validity (MeSH)). Articles were included only if they reported in English on the accuracy of a tools in  
8  
9 164 the inventory against an objective criterion measure of SB, and/or sensitivity to change. Although  
10  
11 165 articles were only included in the review if they assessed accuracy or sensitivity to change, the  
12  
13 166 search terms included a wide range of psychometric properties in order to maximise the chances of  
14  
15 167 finding eligible articles.  
16  
17  
18  
19

168

20  
21  
22 169 Exclusion by title, then abstract, then full-text was conducted by two researchers from a pool of five  
23  
24 170 [PD, EC, CF, SC, CL]. In the case of disagreement, the article was carried forward in to the next  
25  
26 171 round, or at full-text stage a third researcher was consulted to ensure consensus. Data (tool,  
27  
28 172 criterion, population, statistical analysis, accuracy of sedentary behaviour, sensitivity to change of  
29  
30 173 sedentary behaviour) was extracted and quality was assessed independently by two researchers  
31  
32 174 from a pool of three [PD, CF, SC]. Disagreements were resolved by discussion. Quality was assessed  
33  
34 175 using QUALSYST [18], modified to include an additional item for the criterion measure. As per the  
35  
36 176 QUALSYST guidelines, the quality score for the article (range 0-1) was used to identify common  
37  
38 177 methodological strengths and flaws, rather than as an objective representation of high/low quality.  
39  
40 178 Accuracy and sensitivity to change extracted from included articles were reported for tools in  
41  
42 179 relation to the TASST taxonomy.  
43  
44  
45

180

## 181 **RESULTS**

### 182 **Inventory**

183 The systematic inventory identified 37 distinct self-report tools used to measure SB in adults and  
184 older adults, 32 of which were identified in the initial search and used to form the taxonomy (Table

1  
2  
3 185 1). The International Physical Activity Questionnaire (IPAQ) was originally developed with four  
4  
5 186 different versions, which were included separately in the inventory (combinations of the long and  
6  
7 187 short versions, and last seven days and usual week recall). The 45 and Up study asked different  
8  
9 188 questions in its baseline and follow-up questionnaires, which have been included as separate tools.  
10  
11 189 Three tools, termed “modified” versions, were included where questions had been added or  
12  
13 190 modified to the original tool (EPAQ2, NHANES, and IPAQ-L, representing a 5<sup>th</sup> version of the IPAQ in  
14  
15 191 the inventory), and were considered to form a substantially different version. Some tools identified  
16  
17 192 were used in only a single study, and these were included in the inventory, referred to by the study  
18  
19 193 name. The 32 tools in the original inventory comprised of 141 individual questions, consisting of  
20  
21 194 between 1 and 20 questions per tool. An evaluation of the content of these individual items formed  
22  
23 195 the basis of the TASST taxonomy.  
24  
25  
26  
27  
28 196

197 Table 1 Tools measuring SB for population surveillance identified in the inventory

Acronym	Name of Tool/Study	Key reference
45Up-B	45 and Up study, baseline questionnaire	[19]
45Up-F	45 and Up study, follow up questionnaire	[19]
ACS2	American Cancer Society, Cancer Prevention Study cohort II	[20]
ALTS	Australian Leisure Time Sitting questionnaire	[21]
AusDiab	The Australian Diabetes Obesity and Lifestyle study	[22]
CCHS	Canadian Community Health Survey	[23]
CFS	Canadian Fitness Survey	[24]
CHAMPS	Community Health Activities Model Program for Seniors physical activity questionnaire	[15]
ELSA	English Longitudinal Study of Ageing	[25]
EPAQ2	European Prospective Investigation of Cancer (EPIC)-Norfolk Physical Activity Questionnaire	[26]
mod EQPAQ2	modified version of the EPIC-Norfolk Physical Activity Questionnaire	[27]
GPAQ	Global Physical Activity Questionnaire	[28]
HSE	Health Survey for England	[29]
HUNT3	Nord-Trøndelag Health Study 3	[30]
IPAQ-L I7d	International Physical Activity Questionnaire, Long version, last 7 days	[31]
IPAQ-L uw	International Physical Activity Questionnaire, Long version, usual week	[31]
mod IPAQ-L	modified version of the International Physical Activity Questionnaire, Long version	[32]
IPAQ-S I7d	International Physical Activity Questionnaire, Short version, last 7 days	[31]
IPAQ-S uw	International Physical Activity Questionnaire, Short version, usual week	[31]
LASA	Longitudinal Aging Study Amsterdam	[33]
MLTPAQ	Minnesota Leisure Time Physical Activity Questionnaire	[34]
MOST	Measuring Older adults' Sedentary Time questionnaire	[35]
NHANES	National Health and Nutrition Examination Survey	[36]
mod NHANES	modified version of the National Health and Nutrition Examination Survey	[37]
NHS2	Nurses Health Survey II	[38]
NIH-AARP DHS	National Institutes of Health – American Association of Retired Persons (NIH-AARP) Diet and Health Survey	[39]
NSWPAS	New South Wales Physical Activity Survey	[40]
PASE	Physical Activity Scale for the Elderly	[41]
PAST	Past-day Adults Sedentary Time questionnaire	[42]
PAST-U*	Past-day Adults Sedentary Time questionnaire – University version	[43]
PCSpa	prospective cohort study (Spain)	[44]
SBQ	Sedentary Behaviour Questionnaire	[45]
SHS	Scottish Health Survey	[46]
SIT-Q*	SIT-Q	[47]
SIT-Q-7d*	past seven day version of the SIT-Q	[48]
STAR-Q*	Sedentary Time and Reporting Questionnaire	[49]
STAQ*	Sedentary, Transportation and Activity Questionnaire	[50]



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

198 Acronym: the commonly used acronym of the tool, or the short identifier adopted for this article Name of Tool: either the  
199 name of the tool, or the name of the single study using these questions/tool. Key reference: references provided here are  
200 not exhaustive, but refer either to an early or well cited description of the tool, or the study in which the tool was used or  
201 adapted. Tools marked with an asterisk (\*) were identified in the updated search, and were not used to create the  
202 taxonomy

For peer review only

1  
2  
3 203 **T**Axonomy for Self-report Sedentary behaviour Tools (TASST)  
4

5  
6 204 The taxonomy derived from the inventory of self-report tools to measure SB (Figure 1) comprises of  
7  
8 205 four domains, which characterise different aspects of the tool: type of assessment, recall period,  
9  
10 206 temporal unit, and assessment period. All four aspects are required to describe the tool. Within  
11  
12 207 each aspect, the taxonomy functions as a tree, meaning you can identify a single end point (taxon)  
13  
14 208 which fully describes each question in a tool.  
15  
16

17  
18 209

19  
20 210 The type of assessment domain of the taxonomy covers the way that the outcome of time spent in  
21  
22 211 SB is derived from the tool. Tools can either ask about a single aspect of SB (1.1 single item), or a  
23  
24 212 composite aspect (1.2 composite). Tools using a single item of assessment will generate all of their  
25  
26 213 information about SB within the relevant period of assessment in a single question. That single item  
27  
28 214 can either ask about sitting time directly (1.1.1 direct measure) or it can ask about a single behaviour  
29  
30 215 related to SB which is then used as a proxy measure of SB duration (1.1.2 proxy measure).  
31  
32 216 Composite items of assessment ask multiple questions about several aspects of SB for the same  
33  
34 217 period of assessment. One form of composite item would be to ask about the pattern (i.e.  
35  
36 218 frequency and timing) of SB accumulated throughout the recall period (1.2.1 pattern). However, the  
37  
38 219 most common form of composite item is created as a sum (1.2.2 sum) of the time spent in SB in a  
39  
40 220 range of different activities or situations. The sum can either be formed from questions asking about  
41  
42 221 specific behaviours (1.2.2.1), activities such as TV viewing, hobbies, talking with friends, or they can  
43  
44 222 be based on domains (1.2.2.2), locations or situations where you can sit, such as at home, for  
45  
46 223 transport and at work.  
47  
48

49  
50  
51 224

52  
53  
54 225 The recall period is total time over which the respondent is asked to consider their SB when  
55  
56 226 answering the questions. The recall period can be anchored to the present time in which case it  
57  
58  
59  
60

1  
2  
3 227 refers to a specific length of time prior to now, for example yesterday (2.1 previous day), last week  
4  
5 228 (2.2 previous week), or a longer period such as the last month or year (2.3 longer). The recall period  
6  
7 229 can also be unanchored (2.4), in which case the respondent is not asked about a specific period but  
8  
9 230 is asked about a general period of time, for example asking about SB in a typical week.  
10

11  
12 231

13  
14  
15 232 The temporal unit is the duration within the recall period that a respondent is asked to report their  
16  
17 233 SB for. For example, in the question “on a typical day last week, how long did you sit?” the recall  
18  
19 234 period is the previous week, but the temporal unit is a day. Within the taxonomy, the temporal  
20  
21 235 units may be a day (3.1), a week (3.2) or longer (3.3). Within a particular recall period, it is possible  
22  
23 236 to have any temporal unit that is of identical or shorter duration than the recall period.  
24  
25  
26

27 237

28  
29  
30 238 The period of assessment is completed by identifying any specific restrictions that are placed on the  
31  
32 239 type of temporal unit recalled. The categories within the assessment period domain clarify whether  
33  
34 240 a respondent is asked questions regarding a particular type of day, for example only about week  
35  
36 241 days (4.1), only weekend days (4.2), or is asked about weekdays and weekend days in separate  
37  
38 242 questions (4.3 both). Additionally, the assessment period domain can identify if a respondent is  
39  
40 243 asked about particular sub divisions of the day (4.4) in separate questions, for example time spent  
41  
42 244 sitting before 6pm. The final taxon in the assessment period is termed ‘not defined’ (4.5), this  
43  
44 245 represents the situation where a respondent is asked about all temporal units (e.g. days) within the  
45  
46 246 recall period (e.g. last week) without any specific distinction being made between them. It is a  
47  
48 247 global category, which usually represents a decision not to separate out these categories, as  
49  
50 248 opposed to a failure to define this domain.  
51  
52

53  
54 249

55  
56  
57 250 **Mapping the Inventory on to the Taxonomy**  
58  
59  
60

1  
2  
3 251 The 37 tools identified in the inventory were mapped against the TASST taxonomy (Table 2).  
4  
5 252 Approximately half of the tools in the inventory (n=17) used a single item of assessment, thirteen  
6  
7 253 used a direct measure and seven used a proxy measure. Three tools (45Up-B, AusDiab, NIH-AARP  
8  
9 254 DHS) asked single item questions about both a direct measure and a proxy measure, but not in a  
10  
11 255 manner in which they could be used as a sum, and have therefore been included in the count for  
12  
13 256 both taxa. Proxy measures were predominantly based on TV viewing (n=5). Twenty tools used  
14  
15 257 composite assessment, all of which used a sum as that composite item. The vast majority of sums  
16  
17 258 were formed from questions asking about different behaviours (n=19), with only one sum formed  
18  
19 259 from questions asking about different domains. The tools using a sum of behaviours generally  
20  
21 260 included the common proxy measures of TV viewing (n=19) and computer use (n=17) within the  
22  
23 261 sum. Many tools included questions for behaviours based on leisure pursuits (n=14), in social  
24  
25 262 contexts (n=9), and during transportation (n=13). Often several behaviours of each type were  
26  
27 263 considered in separate questions (e.g. asking about time sitting while reading separately from time  
28  
29 264 spent sitting listening to music). Questions based on time working were included in ten tools, but  
30  
31 265 were explicitly excluded in four tools. Less frequently, tools included questions based on rest (n=5),  
32  
33 266 or used an “other” category to cover circumstances not explicit within the questions (n=7).  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Table 2: Mapping of the tools measuring SB identified in the inventory onto the TASST taxonomy.

Taxonomy Item	N	Tools	Accuracy	Sensitivity to change
<b>1</b>				
<b>Type of Assessment</b>				
1.1	17		<b>Underestimate with large systematic and a random error</b>	+
1.1.1	13	45Up-B; ACS2; AusDiab; CFS; GPAQ; HUNT3; IPAQ-L I7d; IPAQ-L uw; IPAQ-S I7d; IPAQ-S uw; NIH-AARP DHS; PASE; PCSPa		
1.1.2	7	45Up-B; AusDiab; ELSA; MLTPAQ; NIH-AARP DHS; NSWPAS; SHS		
1.2	20			
1.2	0		<b>Smaller systematic error but there is a potential to overestimate</b>	+
1.2.1	0			
1.2.2	19	45Up-F; ALTS; CCHS; CHAMPS; EPAQ2; mod EPAQ2; HSE; mod IPAQ-L; LASA; MOST; NHANES; mod NHANES; PAST; PAST-U; SBQ; SIT-Q; SIT-Q-7d; STAR-Q; STAQ		
1.2.2.1	1	NHS2		
1.2.2.2	1	NHS2		
<b>2</b>				
<b>Recall period</b>				
2.1	2	PAST; PAST-U	+	-
2.2	9	45Up-F; ALTS; AusDiab; IPAQ-L I7d; IPAQ-S I7d; mod IPAQ-L; MOST; PASE; SIT-Q-7d	-	+
2.3	11	ACS2; CCHS; CHAMPS; EPAQ2; mod EPAQ2; HSE; NHANES; NIH-AARP DHS; SIT-Q; STAR-Q; STAQ	-	-
2.4	15	45Up-B; CFS; ELSA; GPAQ; HUNT3; IPAQ-L uw; IPAQ-S uw; LASA; MLTPAQ; mod NHANES; NHS2; NSWPAS; PCSpa; SBQ; SHS		
<b>3</b>				
<b>Temporal Unit</b>				
3.1	32	45Up-B; 45Up-F; ACS2; AusDiab; CFS; ELSA; EPAQ2; mod EPAQ2; GPAQ; HSE; HUNT3; IPAQ-L I7d; IPAQ-L uw; IPAQ-S I7d; IPAQ-S uw; mod IPAQ-L; LASA; MLTPAQ; NHANES; mod NHANES; NIH-AARP DHS; NSWPAS; PASE; PAST; PAST-U; PCSpa; SBQ; SHS; SIT-Q; SIT-Q-7d; STAR-Q; STAQ	+	+
3.2	5	ALTS; CCHS; CHAMPS; MOST; NHS2	-	-
3.3	0		-	-
<b>4</b>				
<b>Assessment Period</b>				
4.1	2	IPAQ-S I7d; IPAQ-S uw	-	+
4.2	0		-	-
4.3	14	45Up-F; AusDiab; ELSA; HSE; IPAQ-L I7d; IPAQ-L uw; mod IPAQ-L; LASA; NSWPAS; PCSpa; SBQ; SHS; SIT-Q-7d; STAQ	+	-
4.4	1	EPAQ2	+	-
4.5	21	45Up-B; ACS2; ALTS; CCHS; CFS; CHAMPS; EPAQ2; mod EPAQ2; GPAQ; HUNT3; MLTPAQ; MOST; NHS2; NHANES; mod NHANES; NIH-AARP DHS; PASE; PAST; PAST-U; SIT-Q; STAR-Q	<b>Better for older adults</b>	+

Full names for the acronyms reported in the Tools column can be found in Table 1. Recommendations in bold are backed by evidence from the systematic review. Recommendations which are not bold are

theoretical but no evidence could be found in the literature. + represents a positive attribute; - a negative attribute.

1  
2  
3 270 A little under half of the tools in the inventory used an unanchored recall period (n=15), nine used a  
4  
5 271 previous week recall period, and eleven used a longer recall period. Only two tools (PAST, PAST-U)  
6  
7 272 in the inventory used a previous day recall period. The majority of tools used a temporal unit of a  
8  
9 273 day (n=32), with five (ALTS, CCHS, CHAMPS, MOST, NHS2) using a temporal unit of a week. A single  
10  
11 274 question within the EPAQ2 questionnaire was based on a temporal unit longer than a week, but the  
12  
13 275 other three questions in that tool were based on a temporal unit of a day. Just over half the tools  
14  
15 276 (n=21) did not define specific days or time periods in their questions, but asked about the temporal  
16  
17 277 unit within the recall period as a single entity. Sixteen tools used questions specifically referring to  
18  
19 278 week or weekend days, fourteen asking about both week and weekend days, while two asked only  
20  
21 279 about week days. Only one tool (EPAQ2) referred to specific sub-divisions of the day in their  
22  
23 280 questions.  
24  
25  
26  
27  
28  
29  
30  
31  
32

### 32 **Systematic search for measurement characteristics**

33 283 The systematic search returned 7,221 references, and after removal of duplicate and assessment  
34  
35 284 against exclusion criteria (>99% agreement between reviewers), a total of 22 studies were included  
36  
37 285 in the review (figure 2, table 3).  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

286 Table 3: Measurement characteristics of tools measuring SB, presented by tool and taxon

Tool	Taxon (refer to figure 1)	N	Population (Country)	Criterion measure (definition of SB)	QUALSYST Score	Agreement (hours/day) tool - criterion [limit of agreement]	Sensitivity to change	Ref
IPAQ-Long 17d	1.1.1/2.2/3.1/4.3	1508	A & OA (Greenland)	actiHeart (<1.5MET)	0.67	-3.0 [not reported] for adults -6.0 [not reported] for older adults	--	[51]
		542	A (Netherlands)	Actigraph (<100 count/min)	0.78	-1.6 [-6.4 3.2]	--	[52]
		980	A (Sweden)	Actigraph (<100 count/min)	0.67	+2.2 [-4.5 9.5]	--	[53]
		69	A (UK)	activPAL (sitting/lying postures)	0.78	-2.2 [-7.2 3.7]	--	[16]
		317	A (Chile)	Actigraph (<100 count/min)	0.78	-1.1 [-3.8 1.5]	--	[54]
		346	A & OA (Switzerland)	ActiGraph (<150 count/min)	0.78	-3.8 [-9.3 1.7]	--	[55]
IPAQ-Short 17d	1.1.1/2.2/3.1/4.1	1751	A & OA (Norway)	Actigraph (<100 count/min)	0.67	-1.8 [not reported] for adults +3.5 [not reported] for older adults	--	[56]
		144	A (Nigeria)	Actigraph (<100 count/min)	0.78	-3.0 [-8.5 2.5]	--	[57]
		54	OA (Sweden)	Actigraph (<100 count/min)	0.56	-1.5 [not reported]	--	[58]
		127	OA (USA)	Actigraph (<50 count/min)	0.72	-4.4 [-10.0 -1.4]	--	[59]
		50	A & OA (UK)	Actigraph (<50 count/min)	0.72	-0.5 [-1.9 0.8]	--	[60]
GPAQ	1.1.1/2.4/3.1/4.5	62	A (Saudi Arabia)	Actigraph (<100 count/min)	0.67	-3.3 [-9.7 3.1]	--	[61]
CHAMPS	1.2.2.1/2.3/3.2/4.5	870	OA (USA)	Actigraph (<100 count/min)	0.72	-6.8 [-10.6 2.4]	--	[15]
		58	OA (USA)	Actigraph (<100 count/min)	0.72	-5.2 [not reported]	--	[62]

LASA	1.2.2.1/2.4/3.1/4.3	83	OA (Netherlands)	Actigraph (<100 count/min)	0.78	+ 0.2 for 10 item -2.1 [-7.4 3.3] for 6 item	--	[33]
STAQ		88	A (France)	Actigraph (<150 count/min)	0.72	-2.4 [-6.2 4.9]	--	[50]
PAST	1.2.2.1/2.1/3.1/4.5	90	A (Australia)	activPAL (sitting/lying postures)	0.72	-1.0 [- 5.6 3.8]	t-test was inconclusive	[42]
PAST-U		57	A (Australia)	activPAL (sitting/lying postures)	0.78	0.1 [-3.9 4.1]	--	[43]
SIT-Q-7d	1.2.2.1/2.2/3.1/4.3	51	A (Belgium)	activPAL (sitting/lying postures)	0.72	1.0 [-4.8 8.2] for Belgian sample	--	[48]
		402	A (UK)	actiHeart (<1.5MET)		0.4 [-6.9 8.6] for UK sample		
		33 & 33	A & OA (Belgium)	activPAL (sitting/lying postures)	0.83	2.3 [only reported as a %] 0.3 [-8.9 0.7] for older adults	--	[63]
		442	OA (Belgium)	Actigraph (<100 count/min)	0.83	1.36 [-6.0 3.3]	--	[64]
MOST	1.2.2.1/2.2/3.2/4.5	48	OA (Australia)	Actigraph (<100 count/min)	0.67	-3.6 [-7.4 -0.2]	Guyatt Index 0.39 (0.47 for Actigraph)	[35]

287 A: adults; N: number of participants; OA: older adults; Ref: reference; UK: United Kingdom; USA: United States of America. For tool acronyms see table 1.



1  
2  
3 288 Criterion measure  
4  
5

6 289 None of the studies tested the accuracy of the tool against direct observation. Only five studies  
7  
8 290 [16,42,43,48,63] used a postural sensor that actually measures sitting time objectively (activPAL),  
9  
10 291 the other seventeen used an accelerometer built to measure low movement as a criterion measure  
11  
12 292 (ActiGraph, actiHeart).  
13

14  
15  
16 293  
17

18 294 Statistical analysis  
19

20  
21 295 Accuracy and Limits of Agreement were usually derived from Bland and Altman plots. Sensitivity to  
22  
23 296 change was defined differently in the two articles which reported this measurement characteristic;  
24  
25 297 one used t-test statistics [42], one used the Guyatt Index [35].  
26  
27

28  
29 298  
30

31 299 Study Quality  
32

33  
34 300 Studies which scored highly for quality tended to be purposefully designed to test measurement  
35  
36 301 characteristics, rather than secondary analysis of data collected for another purpose. The most  
37  
38 302 common loss of quality was due to the use of accelerometers which assess low movement (e.g.  
39  
40 303 ActiGraph) as a criterion measure, as this does not measure the primary aspect of the definition of  
41  
42 304 SB (i.e. posture). Another issue which lowered quality was the manipulation of the criterion measure  
43  
44 305 without clear justification. For example, some studies manipulated the count threshold (used to  
45  
46 306 define SB) or included only SB bouts longer than a particular duration without justification or solid  
47  
48 307 rationale.  
49

50  
51  
52 308  
53

54  
55 309 Tools and measurement characteristics  
56  
57  
58  
59  
60

1  
2  
3 310 Table 3 summarises the results reported by these studies, arranged per measurement tool and  
4  
5 311 mapped against the relevant taxon. Very few of the existing tools to measure SB using self-report  
6  
7 312 have actually been investigated for these measurement characteristics. Accuracy has been reported  
8  
9 313 for ten out of the 37 tools identified in the inventory (IPAQ-L 17d, IPAQ-S 17d, GPAQ, MOST, CHAMPS,  
10  
11 314 LASA, PAST, PAST-U, STAQ, SIT-Q-7d). The most tested tools were the IPAQ in its long form, last  
12  
13 315 seven days [16, 51-55] and short form, last seven days.[56-60]. The SIT-Q-7d was tested in three  
14  
15 316 studies [48, 63-64], and the CHAMPS was investigated in two studies [15, 62]. Information for other  
16  
17 317 tools; GPAQ [61], LASA [33], MOST [35], PAST [42], PAST-U [43], and STAQ [50], come from single  
18  
19 318 studies. Reports of sensitivity to change are only available for two tools; MOST [35] and PAST [42].  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31

320 Taxa tested

321 The literature provides measurement characteristics information for eight distinct full taxa:

322 1.1.1/2.2/3.1/4.3 with six studies on IPAQ-L uw;

323 1.1.1/2.2/3.1/4.1 with five studies on IPAQ-S uw;

324 1.1.1/2.4/3.1/4.5 with one study on GPAQ;

325 1.2.2.1/2.1/3.1/4.5 with one study on PAST and one study on PAST-U;

326 1.2.2.1/2.2/3.1/4.3 with three studies on SIT-Q-7d;

327 1.2.2.1/2.2/3.2/4.5 with one study on MOST;

328 1.2.2.1/2.3/3.2/4.5 with two studies on CHAMPS; and

329 1.2.2.1/2.4/3.1/4.3 with one study on LASA and one study on STAQ.

330

1  
2  
3 331 .  
4  
5  
6 332 For the assessment type, there is information for direct measures via single item (1.1.1, twelve  
7  
8 333 studies) and for composite sums of behaviours (1.2.2.1, ten studies). However, there is no  
9  
10 334 information for direct proxy measures (1.1.2). For recall period, there is information on all four  
11  
12 335 possible categories (2.1 previous day, two studies; 2.2 previous week, sixteen studies; 2.3 longer,  
13  
14 336 two studies; and 2.4 unanchored, two studies). The unanchored recall period (2.4), used by 40% of  
15  
16 337 the tools in the inventory, is particularly under-represented with only two studies in the validation  
17  
18 338 review. For temporal scale there is mostly information for assessment at day scale (3.1, twenty  
19  
20 339 studies) and only three studies for the temporal scale of a week (3.2). This is broadly representative  
21  
22 340 of usage by tools in the inventory. For assessment period there is information for weekdays only  
23  
24 341 (4.1, five studies) or both weekdays and weekend days (4.3, eleven studies) and for tools with the  
25  
26 342 assessment period not defined (4.5, six studies). The assessment period not defined taxon (4.5),  
27  
28 343 used by over half the tools in the inventory, is under-represented by these validation studies.  
29  
30  
31  
32  
33 344

#### 345 Accuracy

346 Information for taxon 1.1.1/2.2/3.1/4.3 (IPAQ-L-I7d) is not equivocal. The majority of studies  
347 reported a large underestimation of total SB time ranging from 1.1 hours in adults [54] to 6 hours in  
348 older adults [51]. One study reported that tools in this taxon overestimate total SB time by 2.2 hours  
349 in adults [53]. It is clear that the systematic error on estimates of total SB time using tools from this  
350 taxon is likely to be very large (several hours/day). The random error is also likely to be very large as  
351 the Limits of Agreement reported were consistently very large. Information for taxon  
352 1.1.1/2.2/3.1/4.1 (IPAQ-S-I7d) is a little more consistent for adults. Tools in this taxon seem to  
353 underestimate total SB time by 1.5 to 3 hours in adults. However, in older adults this was less clear

1  
2  
3 354 with reports of underestimation by 4.4 hours [59] and overestimation by 3.5 hours [56]. In both  
4  
5 355 populations the error and Limits of Agreement were large, but not as large as for the previous taxon.  
6  
7

8 356  
9

10  
11 357 Although not entirely consistent, tools reporting information from a single item as a direct measure  
12  
13 358 of sitting (taxon 1.1.1) tended to underestimate sitting, with underestimation ranging from -0.5 [60]  
14  
15 359 to -6.0 [51] hours per day. Within those tools, the IPAQ-S-17d (reporting only for week days in the  
16  
17 360 past week, taxa 2.2 & 4.1), tended to have better agreement than the IPAQ-L-17d (reporting for both  
18  
19 361 week and weekend days in the past week, taxa 2.2 & 4.3), and the GPAQ (reporting over a longer  
20  
21 362 recall period with the assessment period not defined, taxa 2.4 & 4.5). Tools reporting on a sum of  
22  
23 363 behaviours (taxon 1.2.2.1), were more likely to overestimate sitting than for the single item direct  
24  
25 364 measure (taxon 1.1.1). Tools which reported on a sum of behaviours over the past day or past week  
26  
27 365 (taxa 1.2.2.1 & 2.1 or 2.2), tended to have the closest agreement with objective criterion measures,  
28  
29 366 with most studies reporting agreement between -1.0 and +2.3 hours per day. Tools which reported  
30  
31 367 sum of behaviours over a longer (taxon 2.3) or unanchored (taxon 2.4) recall period or which had a  
32  
33 368 temporal unit of a week (taxon 3.2) reported larger underestimates (-2.1 to -6.8 hours/day). In  
34  
35 369 particular the CHAMPS tool, reporting both for a recall period of a year (taxon 2.3) with a temporal  
36  
37 370 unit of a week (taxon 3.2), had the largest differences for any tool. However, there were only a few  
38  
39 371 studies reporting on these aspects, and such conclusions are necessarily tentative. Regardless of  
40  
41 372 level of agreement, limits of agreement were large for all tools.  
42  
43  
44  
45

46 373  
47

48  
49 374 Sensitivity to change  
50

51  
52 375 There is almost no information about sensitivity to change. The two studies that assessed sensitivity  
53  
54 376 to change [35, 42] provided little tangible information. The results were either inconclusive [42], or  
55  
56 377 reported the Guyatt index against a criterion measure which does not measure sitting [35]. While  
57  
58  
59  
60

1  
2  
3 378 the latter provided some indication that the tools' sensitivity to change was similar to that of an  
4  
5 379 objective measure of low movement it does not give a clear indication as to whether it is sensitive to  
6  
7 380 a change in total SB time. Neither of these studies reported the minimal detectable change [65], a  
8  
9 381 metric which provides an easily interpretable value of the capacity of a tool to detect a change.  
10  
11

12 382

### 13 14 15 383 **DISCUSSION**

16  
17  
18 384 A taxonomy (TASST) for the systematic description and comparison of self-reported measures of SB  
19  
20 385 has been established. TASST provides a rigorous framework for informed choice, development and  
21  
22 386 evaluation of self-report tools. This framework has been used to review the measurement  
23  
24 387 characteristics of existing tools in order to identify the optimum tool for population surveillance. The  
25  
26 388 available evidence about measurement characteristics essential for population surveillance, namely  
27  
28 389 accuracy and responsiveness to change, was insufficient to ascertain which tool currently used in  
29  
30 390 practice is best. Accuracy was poor for all existing tools, with both under and over estimation of  
31  
32 391 total time spent in SB and large limits of agreement. In addition, there is a complete lack of evidence  
33  
34 392 about their sensitivity to change. Mapping available evidence onto the TASST framework has  
35  
36 393 enabled informed recommendations to be made about the promising features for a surveillance  
37  
38 394 tool, and identification of the aspects on which future research and development of SB surveillance  
39  
40 395 tools should focus.  
41  
42  
43  
44

45 396

46  
47  
48 397 The use of a coherent and robust taxonomy (TASST) to systematically evaluate and compare the  
49  
50 398 characteristics of measurement tools is the main strength of this study. However, in terms of  
51  
52 399 accuracy and sensitivity to change, the current published evidence does not cover the entire  
53  
54 400 taxonomy. Consequently, at present, only tentative recommendations can be provided. The  
55  
56 401 taxonomy can be used, however, to identify gaps in current research and provide focussed guidance  
57  
58  
59  
60

1  
2  
3 402 for future research and development. During the development of TASST, self-report tools which  
4  
5 403 aimed to measure SB in specific populations (e.g. children, those with arthritis) or specialised  
6  
7 404 contexts (e.g. workplace) were not considered. However, TASST is a generic framework, so tools  
8  
9 405 specific to these populations may already be fully described by the taxonomy. For example, a  
10  
11 406 question asking about time spent sitting at school which is specific to children, would be covered  
12  
13 407 under the sub-division of the day assessment period (taxon 4.4). Another consequence of the  
14  
15 408 exclusion criteria is that evidence on accuracy and sensitivity to change of tools specific to these  
16  
17 409 populations was not mapped on the taxonomy. Therefore, the conclusions drawn from the  
18  
19 410 measurement characteristics in this study are only valid for adults and older adults. In addition, this  
20  
21 411 study has the general limitations common to most systematic reviews, i.e. included articles were  
22  
23 412 restricted to those written in English, articles and tools published after the date of search were not  
24  
25 413 included, and any relevant articles not identified during the search will have been excluded.  
26  
27  
28  
29  
30  
31

414

32  
33 415 The current study is the first to clearly define and focus on the measurement characteristics required  
34  
35 416 for population surveillance (accuracy and sensitivity to change). There is only one other systematic  
36  
37 417 review reporting on the measurement characteristics of self-report tools to measure SB [13], which  
38  
39 418 concentrated on validity (assessed through rank correlation) and reliability, which are the  
40  
41 419 measurement characteristics relevant to establishing association between SB and health. In  
42  
43 420 agreement with the previous review, we found that the major flaw of most validation studies was  
44  
45 421 the use of an inadequate criterion measure. The choice of criterion measure depends on the  
46  
47 422 purpose of the tool. While direct observation should be considered the gold standard, if the purpose  
48  
49 423 is to assess total sedentary time, then accurate postural sensors should be adequate (e.g. activPAL).  
50  
51 424 In this review, only five out of 22 studies used an adequate criterion measure. Instead, many studies  
52  
53 425 used an accelerometer which measures low levels of movement at the hip (e.g. ActiGraph) as a  
54  
55 426 criterion measure, but such tools do not measure SB directly and can misclassify standing as sitting  
56  
57  
58  
59  
60

1  
2  
3 427 [12]. Although it is possible that criterion measure may have provided a confounding effect on  
4  
5 428 agreement (e.g. tools assessing previous day recall period (taxon 2.1, PAST, PAST-U) were only  
6  
7 429 assessed against the activPAL, no clear trend towards better or worse agreement with a particular  
8  
9 430 type of criterion measure or ActiGraph cut-off was apparent.  
10

11  
12 431

13  
14  
15 432 Despite the incomplete nature of the evidence, TASST enables the identification of desirable  
16  
17 433 characteristics of self-report tools to measure SB when used for population surveillance. Firstly,  
18  
19 434 tools assessing total SB time as a sum of behaviours (taxon 1.2.2.1; CHAMPS, LASA, MOST, PAST,  
20  
21 435 PAST-U, SIT-Q-7d, STAQ) provided better accuracy than single item direct measurement (taxon 1.1.1;  
22  
23 436 IPAQ-L-I7d, IPAQ-S-I7d and GPAQ) tools, especially when comparing tools with equivalent recall  
24  
25 437 periods. However, this will be dependent on the behaviours or domains included within the sum,  
26  
27 438 and whether they are exhaustive, consistent and mutually exclusive. Tools with a non-exhaustive  
28  
29 439 sum will underestimate total time, for example, the Longitudinal Aging Study Amsterdam (LASA),  
30  
31 440 found that a six item sum provided a better correlation with SB across the sample, but that a ten  
32  
33 441 item sum was more accurate [33]. Conversely, tools which contain behaviours which might occur  
34  
35 442 concurrently (such as watching TV and using a tablet computer) may lead to an over-estimate in  
36  
37 443 total SB time [63]. Secondly, tools using a previous day recall period (taxon 2.1, PAST, PAST-U)  
38  
39 444 tended to provide better accuracy than those with longer recall periods (taxa 2.2, 2.3 and 2.4). This  
40  
41 445 corroborates recent research on the validity of computerised survey systems which assess SB using a  
42  
43 446 past-day recall period [17, 66]. However, although tools using previous day recall may more  
44  
45 447 accurate, it is likely that their sensitivity to change will be less good due to the higher underlying  
46  
47 448 variability in daily SB [67].  
48  
49  
50  
51  
52

53 449  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 450 Most tools currently used for population surveillance of SB systematically underestimate the amount  
4  
5 451 of SB by two to four hours per day. Yet, self-report tools are still the most practical and economical  
6  
7 452 means of population surveillance. Therefore, policy makers and clinicians should be aware that  
8  
9 453 reports of population SB time are likely to be grossly underestimated, and should be cognisant of  
10  
11 454 this fact when making decisions on implementing, developing and evaluating policy and public  
12  
13 455 health interventions. In addition, policy makers and clinicians should be cautious in interpreting any  
14  
15 456 reported difference in population SB time as a real change. The dearth of information about  
16  
17 457 sensitivity to change of these tools means that we do not know the magnitude of change required to  
18  
19 458 be certain that a change is real and not background variation. Moving forward, development of  
20  
21 459 national and international surveillance systems should not be undertaken assuming that a tool is  
22  
23 460 adequate because it has been used previously. Instead, investment should be made in research to  
24  
25 461 evaluate the sensitivity to change and accuracy of tools to measure SB, paying attention to the  
26  
27 462 potential trade-off between these two measurement characteristics. Such research should be  
28  
29 463 carefully planned, to ensure that meaningful comparisons are investigated. The TASST taxonomy  
30  
31 464 should be used as a useful framework to facilitate such a systematic approach.  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 466 Acknowledgements  
4  
5

6 467 The named authors present the study on behalf of the Seniors USP Team, which comprises: Dawn A  
7  
8 468 Skelton (PI), Sebastien Chastin, Simon Cox, Elaine Coulter, Iva Čukić, Philippa Dall, Ian Deary, Geoff  
9  
10 469 Der, Manon Dontje, Claire Fitzsimons, Catherine Gale, Jason Gill, Malcolm Granat, Cindy Gray,  
11  
12 470 Carolyn Greig, Elaine Hindle, Karen Laird, Gillian Mead, Nanette Mutrie, Victoria Palmer, Ratko  
13  
14 471 Radakovic, Naveed Sattar, Richard Shaw, John Starr, Sally Stewart, Sally Wyke. The authors would  
15  
16 472 like to thank Calum Leask (CL) for his help in updating both of the searches in November 2016.  
17  
18  
19

20 473  
21

22 474 Competing Interests:  
23  
24

25 475 All authors had financial support from the Medical Research Council for the submitted work; no  
26  
27 476 financial relationships with any organisations that might have an interest in the submitted work in  
28  
29 477 the previous three years; no other relationships or activities that could appear to have influenced  
30  
31 478 the submitted work.  
32  
33

34 479  
35  
36  
37

38 480 Funding:  
39  
40

41 481 The study was performed in preparation for the Seniors USP: Understanding Sedentary Patterns  
42  
43 482 project, which was supported by the Medical Research Council [MR/K025023/1]  
44  
45

46 483  
47  
48

49 484 Authors Contributions:  
50  
51

52 485 SC and DS conceived and supervised the study. PD and SC designed the study. PD, EC, CF and SC  
53  
54 486 collected the data, and performed the systematic review. PD, CF and SC analysed the data for the  
55  
56  
57  
58  
59  
60

1  
2  
3 487 taxonomy. PD and SC analysed the data from the systematic review. All authors interpreted the  
4  
5 488 data and critically revised the manuscript for important intellectual content.  
6

7 489  
8  
9

10 490 Data Sharing: Data can be obtained from the corresponding author on request.  
11

12 491  
13

14  
15 492 Figure Legends  
16

17 493  
18  
19

20  
21 494 Figure 1: TAxonomy of Self-reported Sedentary behaviour Tools (TASST)  
22

23  
24 495  
25

26  
27 496 Figure 2: PRISMA diagram of the validation systematic review  
28

29  
30 497  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 498 References  
4  
5  
6 499  
7  
8  
9 500 1. Kohl HW 3rd, Craig CL, Lambert EV, et al. The pandemic of physical inactivity: global action for  
10  
11 501 public health. *Lancet* 2012;380:294–305.  
12  
13  
14 502 2. Chastin S, Schwartz U, Skelton D. Development of a consensus taxonomy of sedentary behaviors  
15  
16 503 (SIT): report of Delphi round 1. *PLoS One* 2013;8:e82313.  
17  
18  
19 504 3. Sedentary Behaviour Research Network. Standardized use of the terms “sedentary” and  
20  
21 505 “sedentary behaviours”. *Appl Physiol Nutr Metabol* 2012;37:540-2.  
22  
23  
24 506 4. Healy GN, Matthews CE, Dunstan DW, Winkler EAH, Owen N. Sedentary time and cardio-  
25  
26 507 metabolic biomarkers in US adults: NHANES 2003-06. *Eur Heart J* 2011;32:590-7.  
27  
28  
29 508 5. Harvey JA, Chastin SFM, Skelton DA. How sedentary are older people? A systematic review of the  
30  
31 509 amount of sedentary behavior. *J Aging Phys Act* 2015;23:471-87.  
32  
33  
34 510 6. Proper KI, Singh AS, van Mechelen W, Chinapaw MJM. Sedentary behaviors and health outcomes  
35  
36 511 among adults: a systematic review of prospective studies. *Am J Prev Med* 2011;40:174-82.  
37  
38  
39 512 7. Thorp AA, Owen N, Neuhaus M, Dunstan DW. Sedentary behaviors and subsequent health  
40  
41 513 outcomes in adults a systematic review of longitudinal studies, 1996-2011. *Am J Prev Med*  
42  
43 514 2011;41:207-15.  
44  
45  
46 515 8. de Rezende LFM, Rey-López JP, Matsudo VKR, do Carmo Luiz O. Sedentary behavior and health  
47  
48 516 outcomes among older adults: a systematic review. *BMC Public Health* 2014;14:333.  
49  
50  
51  
52 517 9. Dogra S, Stathokostas L. Sedentary behavior and physical activity are independent predictors of  
53  
54 518 successful aging in middle-aged and older adults. *J Aging Res* 2012;2012:190654.  
55  
56  
57  
58  
59  
60

- 1  
2  
3 519 10. Department of Health. Start Active, Stay Active: A report on physical activity from the four home  
4  
5 520 countries' Chief Medical Officers. 2011.  
6  
7  
8 521 11. Guyatt G, Walter S, Norman G. Measuring change over time. assessing the usefulness of  
9  
10 522 evaluative instruments. *J Chron Dis* 1987;40:171-8.  
11  
12  
13 523 12. Kozey-Keadle S, Libertine A, Lyden K, Staudenmayer J, Freedson P. Validation of wearable  
14  
15 524 monitors for assessing sedentary behavior. *Med Sci Sports Exerc* 2011;43:1561-7.  
16  
17  
18 525 13. Healy GN, Clark BK, Winkler EAH, Gardiner PA, Brown WJ, Matthews CE. Measurement of adults'  
19  
20 526 sedentary time in population-based studies. *Am J Prev Med* 2011;41:216-27.  
21  
22  
23 527 14. Bauman A, Ainsworth BE, Sallis JF, et al. The descriptive epidemiology of sitting: a 20 country  
24  
25 528 comparison using the International Physical Activity Questionnaire (IPAQ). *Am J Prev Med*  
26  
27 529 2011;41:228-35.  
28  
29  
30 530 15. Hekler EB, Buman MP, Haskell WL, et al. Reliability and validity of CHAMPS self-reported  
31  
32 531 sedentary-vigorous intensity physical activity in older adults. *J Phys Act Health* 2012;9:225-36.  
33  
34  
35 532 16. Chastin SFM, Culhane B, Dall PM. Questionnaire sitting time (IPAQ) compared to objective  
36  
37 533 measurement via inclinometry (activPAL). *Physiol Meas* 2014;35:2319-28.  
38  
39  
40 534 17. Kozey-Keadle S, Lyden K, Hickey A, et al. Validation of a previous day recall for measuring the  
41  
42 535 location and purpose of active and sedentary behaviours compared to direct observation. *Int J*  
43  
44 536 *Behav Nutr Phys Act* 2014;11:12.  
45  
46  
47  
48 537 18. Kmet LM, Lee RC, Cook LS. Standard quality assessment criteria for evaluating primary research  
49  
50 538 papers from a variety of fields. *Edmonton: Alberta Heritage Foundation for Medical Research*  
51  
52 539 2004;1-22.  
53  
54  
55 540 19. Banks E, Jorm L, Rogers K, Clements M, Bauman A. Screen-time, obesity, ageing and disability:  
56  
57 541 findings from 91 266 participants in the 45 and Up Study. *Public Health Nutr* 2011;14:34-43.  
58  
59  
60

- 1  
2  
3 542 20 Patel AV, Bernstein L, Deka A, et al. Leisure time spent sitting in relation to total mortality in a  
4  
5 543 prospective cohort of US adults. *Am J Epidemiol* 2010;172:419-429.  
6  
7  
8 544 21. Salmon J, Owen N, Crawford D, Bauman A, Sallis JF. Physical activity and sedentary behavior: a  
9  
10 545 population-based study of barriers, enjoyment, and preference. *Health Psychol* 2003;22:178-88.  
11  
12  
13 546 22. Gardiner PA, Healy GN, Eakin EG, et al. Associations between television viewing time and overall  
14  
15 547 sitting time with the metabolic syndrome in older men and women: The Australian Diabetes Obesity  
16  
17 548 and Lifestyle study. *JAGS* 2011;59:788-96.  
18  
19  
20 549 23. Shields M, Tremblay MS. Sedentary behaviour and obesity. *Health Rep.* 2008;19:19-30.  
21  
22  
23 550 24. Katzmarzyk PT, Church TS, Craig Cl, Bouchard C. Sitting time and mortality from all causes,  
24  
25 551 cardiovascular disease, and cancer. *Med Sci Sports Exerc* 2009;41:998-1005.  
26  
27  
28 552 25. Hamer M, Stamatakis E. Screen-based sedentary behaviour, physical activity, and muscle  
29  
30 553 strength in the English Longitudinal Study of Ageing. *PLoS One* 2013;8:e66222.  
31  
32  
33 554 26. Wareham NJ, Jakes RW, Rennie KL, Mitchell J, Hennings S, Day NE. Validity and repeatability of  
34  
35 555 the EPIC-Norfolk physical activity questionnaire. *Int J Epidemiol* 2002;31:186-74.  
36  
37  
38 556 27. Parsons TJ, Thomas C, Power C. Estimated activity patterns in British 45 year olds: cross-  
39  
40 557 sectional findings from the 1958 British birth cohort. *Eur J Clin Nutr* 2009;63:978-85.  
41  
42  
43 558 28. Armstrong T, Bull F. Development of the World Health Organization Global Physical Activity  
44  
45 559 Questionnaire (GPAQ). *J Public Health* 2006;14:66-70.  
46  
47  
48 560 29. Stamatakis E, Davis M, Stathi A, Hamer M. Associations between multiple indicators of  
49  
50 561 objectively-measured and self-reported sedentary behaviour and cardiometabolic risk in older  
51  
52 562 adults. *Prev Med* 2012;54:82-7.  
53  
54  
55  
56  
57  
58  
59  
60

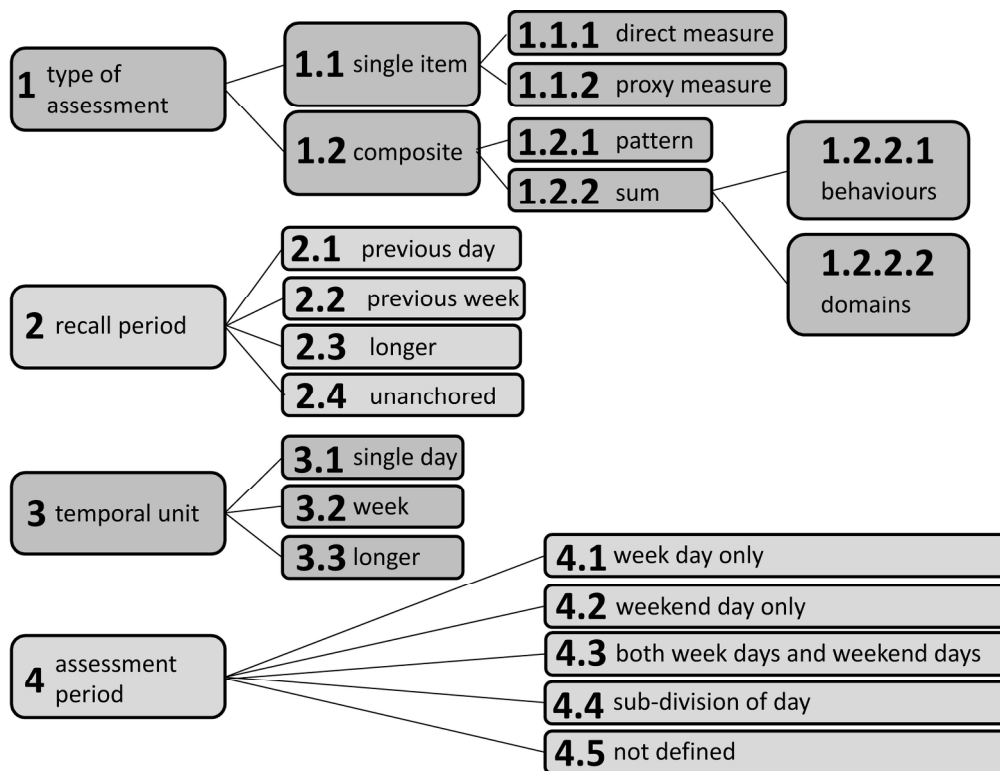
- 1  
2  
3 563 30. Chau JY, Grunseit A, Midthjell K, et al. Sedentary behaviour and risk of mortality from all-causes  
4  
5 564 and cardiometabolic diseases in adults: evidence from the HUNT3 population cohort. *Br J Sports*  
6  
7 565 *Med* 2015;49:737-42.  
8  
9  
10 566 31. Craig CL, Marshall AL, Sjöström M, et al. International Physical Activity Questionnaire: 12-country  
11  
12 567 reliability and validity. *Med Sci Sports Exerc* 2003;35:1381-95.  
13  
14  
15 568 32. Proper KI, Cerin E, Brown WJ, Owen N. Sitting time and socio-economic differences in overweight  
16  
17 569 and obesity. *Int J Obes* 2007;31:169-76.  
18  
19  
20 570 33. Visser M, Koster A. Development of a questionnaire to assess sedentary time in older persons – a  
21  
22 571 comparative study using accelerometry. *BMC Geriatr* 2013;13:80.  
23  
24  
25 572 34. Burazeri G, Goda A, Kark JD. Television viewing, leisure-time exercise and acute coronary  
26  
27 573 syndrome in transitional Albania. *Prev Med* 2008;47:112-5.  
28  
29  
30 574 35. Gardiner PA, Clark BK, Healy GN, Eakin EG, Winkler EAH, Owen N. Measuring older adults'  
31  
32 575 sedentary time: reliability, validity, and responsiveness. *Med Sci Sports Exerc* 2011;43: 2127-33.  
33  
34  
35 576 36. Ford ES. Combined television viewing and computer use and mortality from all-causes and  
36  
37 577 diseases of the circulatory system among adults in the United States. *BMC Public Health* 2012;12:70  
38  
39  
40 578 37. Evenson KR, McGinn AP. Test-retest reliability of adult surveillance measures for physical activity  
41  
42 579 and inactivity. *Am J Prev Med* 2005;28:470-8.  
43  
44  
45 580 38. Wolf AM, Hunter DJ, Colditz GA, et al. Reproducibility and validity of a self-administered physical  
46  
47 581 activity questionnaire. *Int J Epidemiol* 1994;23:991-99.  
48  
49  
50 582 39. Matthews CE, George SM, Moore SC, et al. Amount of time spent in sedentary behaviors and  
51  
52 583 cause-specific mortality in US adults. *Am J Clin Nutr* 2012;ajcn-019620.  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 584 40. Salmon J, Bauman A, Crawford D, Timperio A, Owen N. The association between television  
4  
5 585 viewing and overweight among Australian adults participating in varying levels of leisure-time  
6  
7 586 physical activity. *Int J Obes Relat Metab Disord* 2000;24:600-6.  
8  
9  
10 587 41. Washburn RA, McAuley E, Katula J, Mihalko SL, Boileau RA. The Physical Activity Scale for the  
11  
12 588 Elderly (PASE): evidence for validity. *J Clin Epidemiol* 1999;52:643-51.  
13  
14  
15 589 42. Clark BK, Winkler E, Healy GN, et al. Adults' past-day recall of sedentary time: reliability, validity  
16  
17 590 and responsiveness. *Med Sci Sports Exerc* 2013;45:1198-207.  
18  
19  
20 591 43. Clark BK, Pavey TG, Lim RF, Gomersall SR, Brown WJ. Past-day recall of sedentary time: validity  
21  
22 592 of a self-reported measure of sedentary time in a university population. *J Sci Med Sport* 2016;19:237-  
23  
24 593 41.  
25  
26  
27 594 44. Balboa-Castillo T, León-Muñoz LM, Graciani A, Rodríguez-Artalejo F, Guallar-Castillón P.  
28  
29 595 Longitudinal association of physical activity and sedentary behavior during leisure time with health-  
30  
31 596 related quality of life in community-dwelling older adults. *Health Qual Life Outcomes* 2011;9:41-47.  
32  
33  
34 597 45. Rosenberg DE, Norman GK, Wagner N, Patrick K, Calfas KJ, Sallis JF. Reliability and validity of the  
35  
36 598 Sedentary Behaviour Questionnaire (SBQ) for adults. *J Phys Act Health* 2010;7:697-705.  
37  
38  
39 599 46. Stamatakis E, Hamer M, Dunstan DW. Screen-based entertainment time, all-cause mortality, and  
40  
41 600 cardiovascular events. *J Am Coll Cardiol* 2011;57:292-9.  
42  
43  
44 601 47. Lynch BM, Friedenreich CM, KHandwala F, Liu A, Nicholas J, Csizmadia I. Development and testing  
45  
46 602 of a past year measure of sedentary behaviour: the SIT-Q. *BMC Public Health*; 2014;14:899.  
47  
48  
49 603 48. Wijndaele K, de Bourdeaudhuij I, Godino JG, et al. Reliability and validity of a domain-specific  
50  
51 604 last 7-d sedentary time questionnaire. *Med Sci Sports Exerc* 2014;46:1248-60.  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 605 49. Csizmadi I, Neilson HK, Kopciuk KA, et al. The Sedentary Time and Activity reporting  
4  
5 606 Questionnaire (STAR-Q): reliability and validity against doubly labelled water and 7-day activity  
6  
7 607 diaries. *Am J Epidemiol* 2014;80:424-35.  
8  
9  
10 608 50. Mensah K, Maire A, Oppert J-M, et al. Assessment of sedentary behaviours and transport-  
11  
12 609 related activities by questionnaire: a validation study. *BMC Public Health* 2016;16:753.  
13  
14  
15 610 51. Dahl-Pederson KI, Hansen AW, Bjerregaard P, Jørgensen ME, Brage S. Validity of the  
16  
17 611 international physical activity questionnaire in the arctic. *Med Sci Sports Exerc* 2013;45:728-36.  
18  
19  
20 612 52. van Dyck D, Cardon G, Deforche B, de Bourdeaudhuij I. IPAQ interview version: convergent  
21  
22 613 validity with accelerometers and comparison of physical activity and sedentary time levels with the  
23  
24 614 self-administered version. *J Sports Med Phys Fitness* 2014;55:776-86.  
25  
26  
27  
28 615 53. Hagstromer M, Ainsworth BE, Sjostrom M. Comparison of a subjective and an objective measure  
29  
30 616 of physical activity in a population sample. *J Phys Act Health* 2010;7:541-50.  
31  
32  
33 617 54. Celis-Morales CA, Perez-Bravo F, Ibanez L, Salas C, Bailey MES, Gill JMR. Objective vs. self-  
34  
35 618 reported physical activity and sedentary time: effects of measurement method on relationships with  
36  
37 619 risk biomarkers. *PLoS ONE* 2012;7:e36345.  
38  
39  
40 620 55. Wanner M, Probst-Hensch N, Kriemler S, Meier F, Autenrieth C, Martin BW. Validation of the  
41  
42 621 long international physical activity questionnaire: influence of age and language region. *Prev Med*  
43  
44 622 *Rep* 2016;3:250-6.  
45  
46  
47 623 56. Dyrstad SM, Hansen BH, Holme IM, Anderssen SA. Comparison of self-reported versus  
48  
49 624 accelerometer-measured physical activity. *Med Sci Sports Exerc* 2014;46:99-106.  
50  
51  
52 625 57. Oyeyemi AL, Umar M, Oguiche F, Aliyu SU, Oyeyemi AY. Accelerometer-determined physical  
53  
54 626 activity and its comparison with the International Physical Activity Questionnaire in a sample of  
55  
56 627 Nigerian adults. *PLoS ONE* 2014;9:e87233.  
57  
58  
59  
60



- 1  
2  
3 628 58. Hurtig-Wennlöf A, Hagströmer M, Olsson L. The International Physical Activity Questionnaire  
4  
5 629 modified for the elderly: aspects of validity and feasibility. *Public Health Nutr* 2010;13:1847-54.  
6  
7  
8 630 59. Grimm EK, Swartz AM, Hart T, Miller NE, Strath SJ. Comparison of the IPAQ-short form and  
9  
10 631 accelerometry predictions of physical activity in older adults. *J Aging Phys Act* 2012;20:64-79.  
11  
12  
13 632 60. Curry WB, Thompson JL. Comparability of Accelerometer- and IPAQ-derived physical activity and  
14  
15 633 sedentary time in south Asian women: a cross-sectional study. *Eur J Sport Sci* 2015;15:655-62.  
16  
17  
18 634 61. Alkahtani SA. Convergent validity: agreement between accelerometry and the Global Physical  
19  
20 635 Activity Questionnaire in college-age Saudi men. *BMC Res Notes* 2016;9:436.  
21  
22  
23 636 62. Gennuso KP, Matthews CE, Colbert LH. Reliability and validity of two self-report measures to  
24  
25 637 assess sedentary behavior in older adults. *J Phys Act Health* 2015;12:727-32.  
26  
27  
28 638 63. Busschaert C, de Bourdeaudhuij I, van Holle V, Chastin SFM, Cardon G, de Cocker K. Reliability  
29  
30 639 and validity of three questionnaires measuring context-specific sedentary behaviour and associated  
31  
32 640 correlates in adolescents, adults, and older adults. *Int J Behav Nutr Phys Act* 2015;12:117.  
33  
34  
35 641 64. van Cauwenberg J, van Holle V, de Bourdeaudhuij I, Owen N, Deforche B. Older adults' reporting  
36  
37 642 of specific sedentary behaviours: validity and reliability. *BMC Public Health* 2014;14:734.  
38  
39  
40 643 65. Beaton DE, Bombardier C, Katz JN, Wright JG. A taxonomy for responsiveness. *J Clin Epidemiol*  
41  
42 644 2001;54:1204-71.  
43  
44  
45 645 66. Gomersall SR, Pavey T, Clark BK, Jasman A, Brown WJ. Validity of a self-report recall tool for  
46  
47 646 estimating sedentary behaviour in adults. *J Phys Act Health* 2015;12:1485-91.  
48  
49  
50 647 67. Hart TL, Swartz AM, Cashin SE, Strath SJ. How many days of monitoring predict physical activity  
51  
52 648 and sedentary behaviour in older adults? *Int J Behav Nutr Phys Act* 2011;8:62.  
53  
54  
55  
56  
57  
58  
59  
60



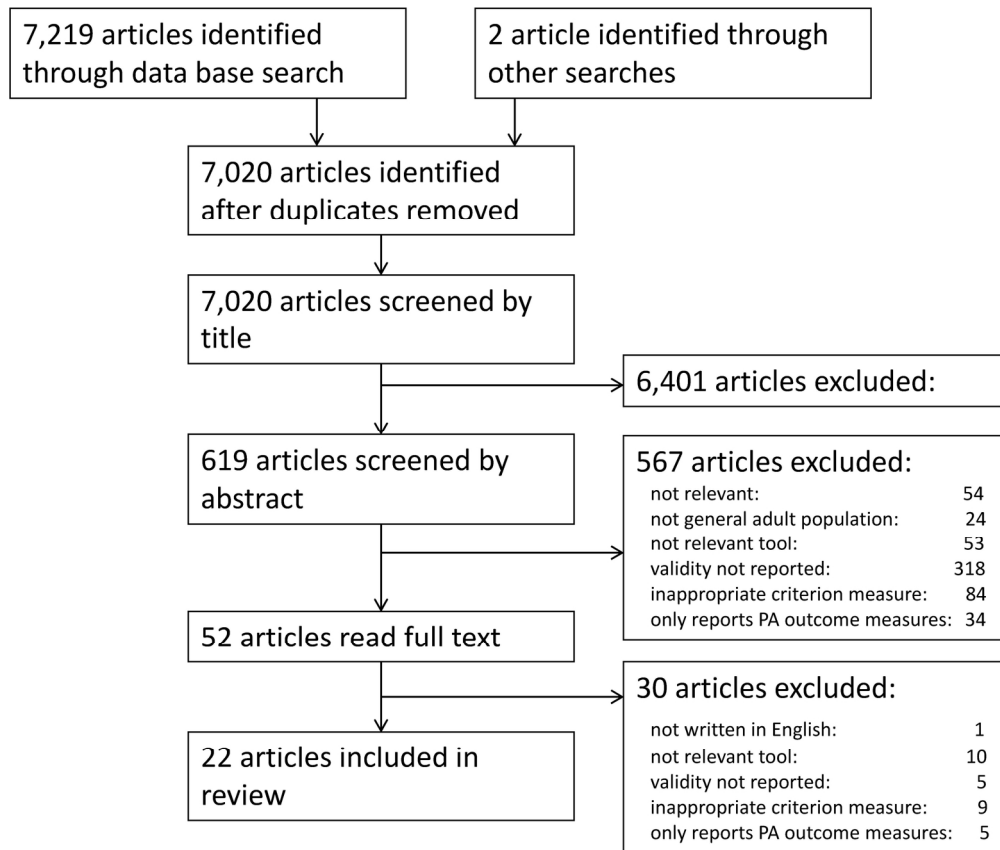
Taxonomy of Self-reported Sedentary behaviour Tools (TASST)

Figure 1

186x142mm (300 x 300 DPI)

ew only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



PRISMA diagram of the validation systematic review

Figure 2

171x144mm (300 x 300 DPI)

For peer review only



# PRISMA 2009 Checklist

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

Section/topic	#	Checklist item	Reported on page #
<b>TITLE</b>			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
<b>ABSTRACT</b>			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
<b>INTRODUCTION</b>			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	3
<b>METHODS</b>			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	4
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	4 & 5
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	4 & 5
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	5
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	5
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	5
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	5
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	Not reporting on trials, so not done



# PRISMA 2009 Checklist

Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	5
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., $I^2$ ) for each meta-analysis.	Not done

Page 1 of 2

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	9
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	Not done
<b>RESULTS</b>			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	8, figure 2
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Table 3
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	8
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Table 3, pages 9 & 10 (not a review of interventions)
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	Not enough data for meta-analysis
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	9
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	Not done
<b>DISCUSSION</b>			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	12
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	11 & 12



# PRISMA 2009 Checklist

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	12
<b>FUNDING</b>			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	14

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: [www.prisma-statement.org](http://www.prisma-statement.org).

Page 2 of 2

# BMJ Open

## The TAxonomy of Self-reported Sedentary behaviour Tools (TASST) framework for development, comparison and evaluation of self-report tools: content analysis and systematic review.

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2016-013844.R2
Article Type:	Research
Date Submitted by the Author:	30-Jan-2017
Complete List of Authors:	Dall, Philippa; Glasgow Caledonian University School of Health and Life Sciences, Department of Psychology, Social Work & Allied Health Sciences Coulter, Elaine; University of Glasgow Fitzsimons, Claire F.; Univ Edinburgh, ISPEHS Skelton, Dawn; Glasgow Caledonian University, School of Health & Life Sciences Chastin, Sebastian; Glasgow Caledonian University
<b>Primary Subject Heading</b>:	Public health
Secondary Subject Heading:	Epidemiology, Research methods, Sports and exercise medicine
Keywords:	sedentary behaviour, sitting, population surveillance, measurement, validation

SCHOLARONE™  
Manuscripts

Only

1  
2  
3 1 **Title:** The TAxonomy of Self-reported Sedentary behaviour Tools (TASST) framework for  
4  
5 2 development, comparison and evaluation of self-report tools: content analysis and systematic  
6  
7 3 review.  
8  
9

10 4  
11  
12  
13 5 **Authors:**

14  
15  
16 6 Dall PM<sup>1</sup>, Coulter EH<sup>2</sup>, Fitzsimons CF<sup>3</sup>, Skelton DA<sup>1</sup>, Chastin SFM<sup>1</sup>, on behalf of the Seniors USP Team.  
17  
18  
19 7

20  
21  
22 8 <sup>1</sup> Institute for Applied Health Research, Glasgow Caledonian University, Cowcaddens Road, Glasgow,  
23  
24 9 G4 0BA

25  
26 10 <sup>2</sup> Department of Nursing and Healthcare, Glasgow University, 59 Oakfield Avenue, Glasgow, G12 8LL  
27

28 11 <sup>3</sup> Sport, Physical Education and Health Sciences Institute, Edinburgh University, St Leonard's Land,  
29  
30 12 Holyrood Road, Edinburgh, EH8 8AQ  
31  
32

33 13  
34  
35  
36 14 corresponding author:

37 15 Dr Philippa Dall

38 16 School of Health & Life Sciences

39 17 Glasgow Caledonian University

40 18 Cowcaddens Road

41 19 Glasgow

42 20 G4 0BA

43 21 philippa.dall@gcu.ac.uk  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54 22

55  
56  
57 23 Word count: 5,107  
58  
59  
60



1  
2  
3 24 **ABSTRACT**  
4  
5

6 25 **Objective:** Sedentary behaviour (SB) has distinct deleterious health outcomes, yet there is no  
7  
8 26 consensus on best practice for measurement. This study aimed to identify the optimal self-report  
9  
10 27 tool for population surveillance of SB, using a systematic framework.  
11

12  
13 28 **Design:** A framework, Taxonomy of Self-report SB Tools (TASST), consisting of four domains (type of  
14  
15 29 assessment, recall period, temporal unit, and assessment period), was developed based on a  
16  
17 30 systematic inventory of existing tools. The inventory was achieved through a systematic review of  
18  
19 31 studies reporting SB and tracing back to the original description. A systematic review of the accuracy  
20  
21 32 and sensitivity to change of these tools was then mapped against TASST domains.  
22  
23

24  
25 33 **Data Sources:** Systematic searches were conducted via EBSCO, reference lists and expert opinion.  
26  
27

28 34 **Eligibility Criteria for selecting studies:** The inventory included tools measuring SB in adults that  
29  
30 35 could be self-completed at one sitting, and excluded tools measuring SB in specific populations or  
31  
32 36 contexts. The systematic review included studies reporting on the accuracy against an objective  
33  
34 37 measure of SB and/or sensitivity to change of a tool in the inventory.  
35  
36

37 38 **Results:** The systematic review initially identified 32 distinct tools (141 questions), which were used  
38  
39 39 to develop the TASST framework. Twenty-two studies evaluated accuracy and/or sensitivity to  
40  
41 40 change representing only 8 taxa. Assessing SB as a sum of behaviours and using a previous day recall  
42  
43 41 were the most promising features of existing tools. Accuracy was poor for all existing tools, with  
44  
45 42 both under and over estimation of SB. There was a lack of evidence about sensitivity to change.  
46  
47

48  
49 43 **Conclusions:** Despite the limited evidence, mapping existing SB tools onto the TASST framework has  
50  
51 44 enabled informed recommendations to be made about the most promising features for a  
52  
53 45 surveillance tool, identified aspects on which future research and development of SB surveillance  
54  
55 46 tools should focus.  
56  
57  
58  
59  
60

1  
2  
3 47 **Systematic Review Registration**  
4  
5

6 48 PROSPERO/CRD42014009851  
7  
8

9 49  
10

11  
12 50 **KEY WORDS:**  
13

14  
15 51 sedentary behaviour; sitting; population surveillance; measurement; validation  
16  
17

18 52  
19

20  
21 53 **STRENGTHS AND LIMITATIONS OF THIS STUDY:**  
22

23  
24 54 • A systematic approach was taken towards classifying self-reported measures of sedentary  
25

26 55 behaviour, allowing a structured approach to measurement in the future  
27

28 56 • An example of use of the framework is presented, mapping accuracy and sensitivity to  
29

30 57 change of self-report sedentary behaviour measures on to the framework  
31

32 58 • Although designed to be generic, the TASST framework was developed excluding tools  
33

34 59 measuring sedentary behaviour in specialised populations and contexts, e.g. children or the  
35

36 60 workplace, and the framework may therefore not cover some aspects of these tools  
37

38 61 • There is the potential for a language bias, as full-text articles not in English were not  
39

40 62 included in the systematic reviews.  
41  
42  
43  
44 63  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 64 **BACKGROUND**  
4  
5

6 65 Physical inactivity is currently at pandemic levels [1] and is a global public health concern. Sedentary  
7  
8 66 behaviour (SB), an umbrella term for all waking time spent in non-exercising sitting or reclining  
9  
10 67 postures [2, 3] such as sitting during work, motorised transport or watching TV, is the largest  
11  
12 68 contributor to inactivity [4,5]. Higher levels of SB have been associated with poor physical and  
13  
14 69 mental health, increased risk of chronic disease and less successful ageing [6-9]. Consequently,  
15  
16 70 several countries, including the UK, have issued recommendations to reduce SB at all ages as part of  
17  
18 71 their national physical activity guidelines [10]. Population surveillance is urgently needed to monitor  
19  
20 72 the impact of such policy, track changes in SB over time, and to evaluate public health interventions  
21  
22 73 targeting SB. In order to provide effective surveillance upon which to base future policy decisions,  
23  
24 74 such surveillance tools should be accurate (provide a true measure of the actual amount of SB in a  
25  
26 75 population) and sensitive to change (provide the true difference in SB between two measurement  
27  
28 76 time points) [11].  
29  
30  
31  
32  
33  
34  
35

36 78 Objective body worn sensors, that measure posture, demonstrate good accuracy for measuring total  
37  
38 79 duration of SB against the gold standard of direct observation [12], but they are expensive and  
39  
40 80 challenging to use for population surveillance. Self-report tools provide a pragmatic choice for  
41  
42 81 population surveillance and have the potential to provide context rich information, useful for  
43  
44 82 intervention development [13]. To date, surveys assessing SB have predominantly used self-report  
45  
46 83 tools [14], which are generally adapted from tools not specifically designed to measure that  
47  
48 84 behaviour (e.g. tools designed to measure physical activity) [15], and which have not been evaluated  
49  
50 85 for population surveillance purposes [14]. No framework currently exists with which to describe and  
51  
52 86 compare SB self-report tools, meaning there is currently no way of systematically selecting an  
53  
54 87 appropriate tool. A previous systematic review of the measurement characteristics of self-report  
55  
56 88 tools measuring SB, reported acceptable to good reliability but low to moderate correlation with a  
57  
58  
59  
60

1  
2  
3 89 (non-gold standard) criterion measure [13]. This suggests that self-report measures of SB are  
4  
5 90 acceptable tools to establish epidemiological evidence of an association between SB and health [13].  
6  
7 91 However, it is possible that the scale of the problem may be vastly underestimated, as differences of  
8  
9 92 2-4 hours per day (approximately 20% of SB) have been reported between self-report and objective  
10  
11 93 tools [16].  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40

94

17 95 The primary aim of this study was to identify, in a systematic manner, the optimal self-report tool to  
18  
19 96 measure SB for use in population surveillance. Although self-report SB tools can and will be used in  
20  
21 97 other areas of research, this study focussed on population surveillance as an area that is crucial to  
22  
23 98 the development of public health policy. To fulfil the primary aim, a framework was created to  
24  
25 99 describe the features of self-report tools measuring SB, the TAXonomy of Self-report Sedentary  
26  
27 100 behaviour Tools (TASST). A systematic inventory of existing self-report tools to measure SB was  
28  
29 101 mapped onto TASST, and the measurement characteristics of these tools, focussing on accuracy and  
30  
31 102 sensitivity to change, were evaluated, with explicit reference to the domains of the taxonomy  
32  
33 103 framework.  
34  
35  
36  
37  
38  
39  
40

104

## 105 **METHODS**

41  
42  
43  
44 106 The study protocol (PROSPERO CRD42014009851), was conducted in three phases. In phase 1 an  
45  
46 107 exhaustive inventory of self-reported tools to measure SB in adults and older adults was established  
47  
48 108 using a structured search protocol. Phase 2 was the development of a taxonomy based on content  
49  
50 109 analysis of the items and questions in the tools. In phase 3, a systematic literature review of the  
51  
52 110 measurement characteristics of the tools in the inventory was conducted and mapped onto the  
53  
54 111 taxonomy.  
55  
56  
57  
58  
59  
60

112

1  
2  
3 113 **Phase 1: Systematic inventory of self-report tools**  
4  
5

6 114 The aim of the systematic inventory was to compile an exhaustive list of self-report tools which  
7  
8 115 could be used to measure SB in adults ( $\geq 18$  years) and older adults ( $\geq 60$  years). Since the aim was to  
9  
10 116 identify tools and not to identify articles, this stage does not have the same methodology as a  
11  
12 117 systematic literature review. A literature search was conducted in October 2013 (updated  
13  
14 118 November 2016), for articles reporting SB as an outcome measure. From this review, a list of self-  
15  
16 119 report tools which measured SB was compiled. References lists were reviewed and experts  
17  
18 120 consulted to identify any additional tools to include in the inventory. The inventory then was  
19  
20 121 consolidated to amalgamate tools referred to by different names, and to trace back to the original  
21  
22 122 version. Articles which added significant new questions to tools were included as a separate tool.  
23  
24 123 We defined significant new questions to be at least one question which added or changed the type  
25  
26 124 of sedentary behaviour or the time period considered by the tool. Changes in phrasing of the  
27  
28 125 question were not considered sufficient to be considered as a separate tool. Tools used in a single  
29  
30 126 study and those without names/acronyms were included as separate tools.  
31  
32  
33  
34  
35  
36  
37

38 127  
39  
40 128 To be included in the inventory, tools had to: be suitable for use for large scale population studies of  
41  
42 129 adults or older adults, including being suitable for self-completion by the respondent at a single  
43  
44 130 point in time (a pragmatic requirement to minimise participant burden); and measure SB or a proxy  
45  
46 131 measure of SB (e.g. TV viewing). Although there is great interest in the sedentary behaviour across  
47  
48 132 many populations and contexts, for pragmatic purposes initial taxonomy development was limited  
49  
50 133 to a core of self-report tools widely applicable to the general adult population. Therefore, tools  
51  
52 134 were excluded from the inventory: if they were designed specifically to assess SB in children or other  
53  
54 135 specialised populations (e.g. medical conditions); if they were designed specifically to assess SB in a  
55  
56 136 specialised context, (e.g. workplace or care settings); if continuous reporting over extended periods  
57  
58 137 of time was required (e.g. diaries or time-use surveys); or if significant interviewer interactions were  
59  
60

1  
2  
3 138 required. Self-report tools that could be administered by telephone or interview were not  
4  
5 139 automatically excluded, however tools such as the PDR (Previous Day Recall) [17], in which the  
6  
7 140 interviewer works through lists of several hundred items, were excluded.  
8  
9  
10  
11 141

## 12 13 142 **Phase 2: Development of a taxonomy**

14  
15  
16 143 Only tools identified in the initial search were used to develop the taxonomy. The original text was  
17  
18 144 extracted for each question relating to SB in each of the self-report tools identified in the inventory.  
19  
20 145 Content analysis was conducted on the text to extract all of the attributes in the questions that were  
21  
22 146 used to describe and constrain what aspect of SB was measured by that question. For example, in  
23  
24 147 the question “During the last 7 days, how much time did you usually spend sitting on a week day?”,  
25  
26 148 attributes extracted relating to the measurement of SB would be “during the last 7 days”, “time  
27  
28 149 spent sitting” and “on a week day”. Attributes were then grouped into mutually exclusive domains  
29  
30 150 covering similar aspects of measurement, and categories within those domains were defined  
31  
32 151 iteratively. A new category was created each time a tool did not fit within an existing category. The  
33  
34 152 full taxonomy was then assembled and streamlined by merging categories with overlapping  
35  
36 153 meaning. Finally, consideration was given to potential future developments of self-report tools to  
37  
38 154 measure SB, such as the growing interest in the pattern of accumulation of sedentary behaviour, by  
39  
40 155 adding any categories to the taxonomy considered useful in the future. The resulting taxonomy was  
41  
42 156 then tested by ensuring that all tools could be classified similarly by two independent researchers  
43  
44 157 and that the taxonomy fully defined the tool.  
45  
46  
47  
48  
49  
50 158

## 51 52 159 **Phase 3: Systematic review of measurement characteristics**

53  
54  
55 160 Finally, a systematic literature search in relevant health databases was conducted in December 2014  
56  
57 161 (updated November 2016) via EBSCO host. The search combined the name of the tool including  
58  
59  
60

1  
2  
3 162 variants and acronyms (except where the acronym was also a common word, e.g. PAST, MOST), with  
4  
5 163 search terms relating to measurement characteristics (valid\* /reliab\* /repons\* /sensitiv\* /calibrat\*  
6  
7 164 /accura\* /agreement /psychometric\* /clinimetric\* /"measurement characteristics" /Reliability and  
8  
9 165 Validity (MeSH)). Articles were included only if they reported in English on the accuracy of a tools in  
10  
11 166 the inventory against an objective criterion measure of SB, and/or sensitivity to change. Although  
12  
13 167 articles were only included in the review if they assessed accuracy or sensitivity to change, the  
14  
15 168 search terms included a wide range of psychometric properties in order to maximise the chances of  
16  
17 169 finding eligible articles.  
18  
19  
20  
21  
22

23  
24 171 Exclusion by title, then abstract, then full-text was conducted by two researchers from a pool of five  
25  
26 172 [PD, EC, CF, SC, CL]. In the case of disagreement, the article was carried forward in to the next  
27  
28 173 round, or at full-text stage a third researcher was consulted to ensure consensus. Data (tool,  
29  
30 174 criterion, population, statistical analysis, accuracy of sedentary behaviour, sensitivity to change of  
31  
32 175 sedentary behaviour) was extracted and quality was assessed independently by two researchers  
33  
34 176 from a pool of three [PD, CF, SC]. Disagreements were resolved by discussion. Quality was assessed  
35  
36 177 using QUALSYST [18], modified to include an additional item for the criterion measure. As per the  
37  
38 178 QUALSYST guidelines, the quality score for the article (range 0-1) was used to identify common  
39  
40 179 methodological strengths and flaws, rather than as an objective representation of high/low quality.  
41  
42 180 Accuracy and sensitivity to change extracted from included articles were reported for tools in  
43  
44 181 relation to the TASST taxonomy.  
45  
46  
47  
48

49 182

## 50 51 183 **RESULTS**

### 52 53 54 184 **Inventory**

55  
56  
57  
58  
59  
60

1  
2  
3 185 The systematic inventory identified 37 distinct self-report tools used to measure SB in adults and  
4  
5 186 older adults, 32 of which were identified in the initial search and used to form the taxonomy (Table  
6  
7 187 1). The International Physical Activity Questionnaire (IPAQ) was originally developed with four  
8  
9 188 different versions, which were included separately in the inventory (combinations of the long and  
10  
11 189 short versions, and last seven days and usual week recall). The 45 and Up study asked different  
12  
13 190 questions in its baseline and follow-up questionnaires, which have been included as separate tools.  
14  
15 191 Three tools, termed “modified” versions, were included where questions had been added or  
16  
17 192 modified to the original tool (EPAQ2, NHANES, and IPAQ-L, representing a 5<sup>th</sup> version of the IPAQ in  
18  
19 193 the inventory), and were considered to form a substantially different version. Some tools identified  
20  
21 194 were used in only a single study, and these were included in the inventory, referred to by the study  
22  
23 195 name. The 32 tools in the original inventory comprised of 141 individual questions, consisting of  
24  
25 196 between 1 and 20 questions per tool. An evaluation of the content of these individual items formed  
26  
27 197 the basis of the TASST taxonomy.  
28  
29  
30  
31  
32 198



199 Table 1 Tools measuring SB for population surveillance identified in the inventory

Acronym	Name of Tool/Study	Key reference
45Up-B	45 and Up study, baseline questionnaire	[19]
45Up-F	45 and Up study, follow up questionnaire	[19]
ACS2	American Cancer Society, Cancer Prevention Study cohort II	[20]
ALTS	Australian Leisure Time Sitting questionnaire	[21]
AusDiab	The Australian Diabetes Obesity and Lifestyle study	[22]
CCHS	Canadian Community Health Survey	[23]
CFS	Canadian Fitness Survey	[24]
CHAMPS	Community Health Activities Model Program for Seniors physical activity questionnaire	[15]
ELSA	English Longitudinal Study of Ageing	[25]
EPAQ2	European Prospective Investigation of Cancer (EPIC)-Norfolk Physical Activity Questionnaire	[26]
mod EQPAQ2	modified version of the EPIC-Norfolk Physical Activity Questionnaire	[27]
GPAQ	Global Physical Activity Questionnaire	[28]
HSE	Health Survey for England	[29]
HUNT3	Nord-Trøndelag Health Study 3	[30]
IPAQ-L I7d	International Physical Activity Questionnaire, Long version, last 7 days	[31]
IPAQ-L uw	International Physical Activity Questionnaire, Long version, usual week	[31]
mod IPAQ-L	modified version of the International Physical Activity Questionnaire, Long version	[32]
IPAQ-S I7d	International Physical Activity Questionnaire, Short version, last 7 days	[31]
IPAQ-S uw	International Physical Activity Questionnaire, Short version, usual week	[31]
LASA	Longitudinal Aging Study Amsterdam	[33]
MLTPAQ	Minnesota Leisure Time Physical Activity Questionnaire	[34]
MOST	Measuring Older adults' Sedentary Time questionnaire	[35]
NHANES	National Health and Nutrition Examination Survey	[36]
mod NHANES	modified version of the National Health and Nutrition Examination Survey	[37]
NHS2	Nurses Health Survey II	[38]
NIH-AARP DHS	National Institutes of Health – American Association of Retired Persons (NIH-AARP) Diet and Health Survey	[39]
NSWPAS	New South Wales Physical Activity Survey	[40]
PASE	Physical Activity Scale for the Elderly	[41]
PAST	Past-day Adults Sedentary Time questionnaire	[42]
PAST-U*	Past-day Adults Sedentary Time questionnaire – University version	[43]
PCSpa	prospective cohort study (Spain)	[44]
SBQ	Sedentary Behaviour Questionnaire	[45]
SHS	Scottish Health Survey	[46]
SIT-Q*	SIT-Q	[47]
SIT-Q-7d*	past seven day version of the SIT-Q	[48]
STAR-Q*	Sedentary Time and Reporting Questionnaire	[49]
STAQ*	Sedentary, Transportation and Activity Questionnaire	[50]

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

200 Acronym: the commonly used acronym of the tool, or the short identifier adopted for this article. Name of Tool: either the  
201 name of the tool, or the name of the single study using these questions/tool. Key reference: references provided here are  
202 not exhaustive, but refer either to an early or well cited description of the tool, or the study in which the tool was used or  
203 adapted. Tools marked with an asterisk (\*) were identified in the updated search, and were not used to create the  
204 taxonomy

For peer review only

1  
2  
3 205 **T**Axonomy for Self-report Sedentary behaviour Tools (TASST)  
4

5  
6 206 The taxonomy derived from the inventory of self-report tools to measure SB (Figure 1) comprises of  
7  
8 207 four domains, which characterise different aspects of the tool: type of assessment, recall period,  
9  
10 208 temporal unit, and assessment period. All four aspects are required to describe the tool. Within  
11  
12 209 each aspect, the taxonomy functions as a tree, meaning you can identify a single end point (taxon)  
13  
14 210 which fully describes each question in a tool.  
15  
16  
17  
18  
19

20  
21 212 The type of assessment domain of the taxonomy covers the way that the outcome of time spent in  
22  
23 213 SB is derived from the tool. Tools can either ask about a single aspect of SB (1.1 single item), or a  
24  
25 214 composite aspect (1.2 composite). Tools using a single item of assessment will generate all of their  
26  
27 215 information about SB within the relevant period of assessment in a single question. That single item  
28  
29 216 can either ask about sitting time directly (1.1.1 direct measure) or it can ask about a single behaviour  
30  
31 217 related to SB which is then used as a proxy measure of SB duration (1.1.2 proxy measure).  
32  
33 218 Composite items of assessment ask multiple questions about several aspects of SB for the same  
34  
35 219 period of assessment. One form of composite item would be to ask about the pattern (i.e.  
36  
37 220 frequency and timing) of SB accumulated throughout the recall period (1.2.1 pattern). However, the  
38  
39 221 most common form of composite item is created as a sum (1.2.2 sum) of the time spent in SB in a  
40  
41 222 range of different activities or situations. The sum can either be formed from questions asking about  
42  
43 223 specific behaviours (1.2.2.1), activities such as TV viewing, hobbies, talking with friends, or they can  
44  
45 224 be based on domains (1.2.2.2), locations or situations where you can sit, such as at home, for  
46  
47 225 transport and at work.  
48  
49

50  
51 226  
52

53  
54  
55 227 The recall period is total time over which the respondent is asked to consider their SB when  
56  
57 228 answering the questions. The recall period can be anchored to the present time in which case it  
58  
59  
60

1  
2  
3 229 refers to a specific length of time prior to now, for example yesterday (2.1 previous day), last week  
4  
5 230 (2.2 previous week), or a longer period such as the last month or year (2.3 longer). The recall period  
6  
7 231 can also be unanchored (2.4), in which case the respondent is not asked about a specific period but  
8  
9 232 is asked about a general period of time, for example asking about SB in a typical week.  
10

11  
12 233

13  
14  
15 234 The temporal unit is the duration within the recall period that a respondent is asked to report their  
16  
17 235 SB for. For example, in the question “on a typical day last week, how long did you sit?” the recall  
18  
19 236 period is the previous week, but the temporal unit is a day. Within the taxonomy, the temporal  
20  
21 237 units may be a day (3.1), a week (3.2) or longer (3.3). Within a particular recall period, it is possible  
22  
23 238 to have any temporal unit that is of identical or shorter duration than the recall period.  
24  
25

26  
27 239

28  
29  
30 240 The period of assessment is completed by identifying any specific restrictions that are placed on the  
31  
32 241 type of temporal unit recalled. The categories within the assessment period domain clarify whether  
33  
34 242 a respondent is asked questions regarding a particular type of day, for example only about week  
35  
36 243 days (4.1), only weekend days (4.2), or is asked about weekdays and weekend days in separate  
37  
38 244 questions (4.3 both). Additionally, the assessment period domain can identify if a respondent is  
39  
40 245 asked about particular sub divisions of the day (4.4) in separate questions, for example time spent  
41  
42 246 sitting before 6pm. The final taxon in the assessment period is termed ‘not defined’ (4.5), this  
43  
44 247 represents the situation where a respondent is asked about all temporal units (e.g. days) within the  
45  
46 248 recall period (e.g. last week) without any specific distinction being made between them. It is a  
47  
48 249 global category, which usually represents a decision not to separate out these categories, as  
49  
50 250 opposed to a failure to define this domain.  
51  
52

53  
54  
55 251

56  
57 252 **Mapping the Inventory on to the Taxonomy**  
58  
59  
60

1  
2  
3 253 The 37 tools identified in the inventory were mapped against the TASST taxonomy (Table 2).  
4  
5 254 Approximately half of the tools in the inventory (n=17) used a single item of assessment, thirteen  
6  
7 255 used a direct measure and seven used a proxy measure. Three tools (45Up-B, AusDiab, NIH-AARP  
8  
9 256 DHS) asked single item questions about both a direct measure and a proxy measure, but not in a  
10  
11 257 manner in which they could be used as a sum, and have therefore been included in the count for  
12  
13 258 both taxa. Proxy measures were predominantly based on TV viewing (n=5). Twenty tools used  
14  
15 259 composite assessment, all of which used a sum as that composite item. The vast majority of sums  
16  
17 260 were formed from questions asking about different behaviours (n=19), with only one sum formed  
18  
19 261 from questions asking about different domains. The tools using a sum of behaviours generally  
20  
21 262 included the common proxy measures of TV viewing (n=19) and computer use (n=17) within the  
22  
23 263 sum. Many tools included questions for behaviours based on leisure pursuits (n=14), in social  
24  
25 264 contexts (n=9), and during transportation (n=13). Often several behaviours of each type were  
26  
27 265 considered in separate questions (e.g. asking about time sitting while reading separately from time  
28  
29 266 spent sitting listening to music). Questions based on time working were included in ten tools, but  
30  
31 267 were explicitly excluded in four tools. Less frequently, tools included questions based on rest (n=5),  
32  
33 268 or used an “other” category to cover circumstances not explicit within the questions (n=7).  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Table 2: Mapping of the tools measuring SB identified in the inventory onto the TASST taxonomy.

Taxonomy Item	N	Tools	Accuracy	Sensitivity to change
<b>1</b>				
<b>Type of Assessment</b>				
1.1	17		<b>Underestimate with large systematic and a random error</b>	+
1.1.1	13	45Up-B; ACS2; AusDiab; CFS; GPAQ; HUNT3; IPAQ-L I7d; IPAQ-L uw; IPAQ-S I7d; IPAQ-S uw; NIH-AARP DHS; PASE; PCSPa		
1.1.2	7	45Up-B; AusDiab; ELSA; MLTPAQ; NIH-AARP DHS; NSWPAS; SHS		
1.2	20		<b>Smaller systematic error but there is a potential to overestimate</b>	+
1.2.1	0			
1.2.2	19	45Up-F; ALTS; CCHS; CHAMPS; EPAQ2; mod EPAQ2; HSE; mod IPAQ-L; LASA; MOST; NHANES; mod NHANES; PAST; PAST-U; SBQ; SIT-Q; SIT-Q-7d; STAR-Q; STAQ		
1.2.2.1	1	NHS2		
1.2.2.2				
<b>2</b>				
<b>Recall period</b>				
2.1	2	PAST; PAST-U	<b>+</b>	-
2.2	9	45Up-F; ALTS; AusDiab; IPAQ-L I7d; IPAQ-S I7d; mod IPAQ-L; MOST; PASE; SIT-Q-7d	-	+
2.3	11	ACS2; CCHS; CHAMPS; EPAQ2; mod EPAQ2; HSE; NHANES; NIH-AARP DHS; SIT-Q; STAR-Q; STAQ	-	-
2.4	15	45Up-B; CFS; ELSA; GPAQ; HUNT3; IPAQ-L uw; IPAQ-S uw; LASA; MLTPAQ; mod NHANES; NHS2; NSWPAS; PCSpa; SBQ; SHS		
<b>3</b>				
<b>Temporal Unit</b>				
3.1	32	45Up-B; 45Up-F; ACS2; AusDiab; CFS; ELSA; EPAQ2; mod EPAQ2; GPAQ; HSE; HUNT3; IPAQ-L I7d; IPAQ-L uw; IPAQ-S I7d; IPAQ-S uw; mod IPAQ-L; LASA; MLTPAQ; NHANES; mod NHANES; NIH-AARP DHS; NSWPAS; PASE; PAST; PAST-U; PCSpa; SBQ; SHS; SIT-Q; SIT-Q-7d; STAR-Q; STAQ	<b>+</b>	+
3.2	5	ALTS; CCHS; CHAMPS; MOST; NHS2	-	-
3.3	0		-	-
<b>4</b>				
<b>Assessment Period</b>				
4.1	2	IPAQ-S I7d; IPAQ-S uw	-	+
4.2	0		-	-
4.3	14	45Up-F; AusDiab; ELSA; HSE; IPAQ-L I7d; IPAQ-L uw; mod IPAQ-L; LASA; NSWPAS; PCSpa; SBQ; SHS; SIT-Q-7d; STAQ	<b>+</b>	-
4.4	1	EPAQ2	<b>+</b>	-
4.5	21	45Up-B; ACS2; ALTS; CCHS; CFS; CHAMPS; EPAQ2; mod EPAQ2; GPAQ; HUNT3; MLTPAQ; MOST; NHS2; NHANES; mod NHANES; NIH-AARP DHS; PASE; PAST; PAST-U; SIT-Q; STAR-Q	<b>Better for older adults</b>	+

Full names for the acronyms reported in the Tools column can be found in Table 1. Recommendations in bold are backed by evidence from the systematic review. Recommendations which are not bold are

theoretical but no evidence could be found in the literature. + represents a positive attribute; - a negative attribute.

1  
2  
3 272 A little under half of the tools in the inventory used an unanchored recall period (n=15), nine used a  
4  
5 273 previous week recall period, and eleven used a longer recall period. Only two tools (PAST, PAST-U)  
6  
7 274 in the inventory used a previous day recall period. The majority of tools used a temporal unit of a  
8  
9 275 day (n=32), with five (ALTS, CCHS, CHAMPS, MOST, NHS2) using a temporal unit of a week. A single  
10  
11 276 question within the EPAQ2 questionnaire was based on a temporal unit longer than a week, but the  
12  
13 277 other three questions in that tool were based on a temporal unit of a day. Just over half the tools  
14  
15 278 (n=21) did not define specific days or time periods in their questions, but asked about the temporal  
16  
17 279 unit within the recall period as a single entity. Sixteen tools used questions specifically referring to  
18  
19 280 week or weekend days, fourteen asking about both week and weekend days, while two asked only  
20  
21 281 about week days. Only one tool (EPAQ2) referred to specific sub-divisions of the day in their  
22  
23 282 questions.  
24  
25  
26  
27  
28  
29  
30  
31 283

#### 32 **284 Systematic search for measurement characteristics**

33 285 The systematic search returned 7,221 references, and after removal of duplicate and assessment  
34  
35 286 against exclusion criteria (>99% agreement between reviewers), a total of 22 studies were included  
36  
37 287 in the review (figure 2, table 3).  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

288 Table 3: Measurement characteristics of tools measuring SB, presented by tool and taxon

Tool	Taxon (refer to figure 1)	N	Population (Country)	Criterion measure (definition of SB)	QUALSYST Score	Agreement (hours/day) tool - criterion [limit of agreement]	Sensitivity to change	Ref
IPAQ-Long 17d	1.1.1/2.2/3.1/4.3	1508	A & OA (Greenland)	actiHeart (<1.5MET)	0.67	-3.0 [not reported] for adults -6.0 [not reported] for older adults	--	[51]
		542	A (Netherlands)	Actigraph (<100 count/min)	0.78	-1.6 [-6.4 3.2]	--	[52]
		980	A (Sweden)	Actigraph (<100 count/min)	0.67	+2.2 [-4.5 9.5]	--	[53]
		69	A (UK)	activPAL (sitting/lying postures)	0.78	-2.2 [-7.2 3.7]	--	[16]
		317	A (Chile)	Actigraph (<100 count/min)	0.78	-1.1 [-3.8 1.5]	--	[54]
		346	A & OA (Switzerland)	ActiGraph (<150 count/min)	0.78	-3.8 [-9.3 1.7]	--	[55]
IPAQ-Short 17d	1.1.1/2.2/3.1/4.1	1751	A & OA (Norway)	Actigraph (<100 count/min)	0.67	-1.8 [not reported] for adults +3.5 [not reported] for older adults	--	[56]
		144	A (Nigeria)	Actigraph (<100 count/min)	0.78	-3.0 [-8.5 2.5]	--	[57]
		54	OA (Sweden)	Actigraph (<100 count/min)	0.56	-1.5 [not reported]	--	[58]
		127	OA (USA)	Actigraph (<50 count/min)	0.72	-4.4 [-10.0 -1.4]	--	[59]
		50	A & OA (UK)	Actigraph (<50 count/min)	0.72	-0.5 [-1.9 0.8]	--	[60]
GPAQ	1.1.1/2.4/3.1/4.5	62	A (Saudi Arabia)	Actigraph (<100 count/min)	0.67	-3.3 [-9.7 3.1]	--	[61]
CHAMPS	1.2.2.1/2.3/3.2/4.5	870	OA (USA)	Actigraph (<100 count/min)	0.72	-6.8 [-10.6 2.4]	--	[15]
		58	OA (USA)	Actigraph (<100 count/min)	0.72	-5.2 [not reported]	--	[62]



LASA	1.2.2.1/2.4/3.1/4.3	83	OA (Netherlands)	Actigraph (<100 count/min)	0.78	+ 0.2 for 10 item -2.1 [-7.4 3.3] for 6 item	--	[33]
STAQ		88	A (France)	Actigraph (<150 count/min)	0.72	-2.4 [-6.2 4.9]	--	[50]
PAST	1.2.2.1/2.1/3.1/4.5	90	A (Australia)	activPAL (sitting/lying postures)	0.72	-1.0 [- 5.6 3.8]	t-test was inconclusive	[42]
PAST-U		57	A (Australia)	activPAL (sitting/lying postures)	0.78	0.1 [-3.9 4.1]	--	[43]
SIT-Q-7d	1.2.2.1/2.2/3.1/4.3	51	A (Belgium)	activPAL (sitting/lying postures)	0.72	1.0 [-4.8 8.2] for Belgian sample	--	[48]
		402	A (UK)	actiHeart (<1.5MET)		0.4 [-6.9 8.6] for UK sample		
		33 & 33	A & OA (Belgium)	activPAL (sitting/lying postures)	0.83	2.3 [only reported as a %] 0.3 [-8.9 0.7] for older adults	--	[63]
		442	OA (Belgium)	Actigraph (<100 count/min)	0.83	1.36 [-6.0 3.3]	--	[64]
MOST	1.2.2.1/2.2/3.2/4.5	48	OA (Australia)	Actigraph (<100 count/min)	0.67	-3.6 [-7.4 -0.2]	Guyatt Index 0.39 (0.47 for Actigraph)	[35]

289 A: adults; N: number of participants; OA: older adults; Ref: reference; UK: United Kingdom; USA: United States of America. For tool acronyms see table 1.

1  
2  
3 290 Criterion measure  
4  
5

6 291 None of the studies tested the accuracy of the tool against direct observation. Only five studies  
7

8 292 [16,42,43,48,63] used a postural sensor that actually measures sitting time objectively (activPAL),  
9

10 293 the other seventeen used an accelerometer built to measure low movement as a criterion measure  
11

12 294 (ActiGraph, actiHeart).  
13  
14

15 295  
16  
17

18 296 Statistical analysis  
19

20  
21 297 Accuracy and Limits of Agreement were usually derived from Bland and Altman plots. Sensitivity to  
22

23 298 change was defined differently in the two articles which reported this measurement characteristic;  
24

25 299 one used t-test statistics [42], one used the Guyatt Index [35].  
26  
27

28 300  
29  
30

31 301 Study Quality  
32

33  
34 302 Studies which scored highly for quality tended to be purposefully designed to test measurement  
35

36 303 characteristics, rather than secondary analysis of data collected for another purpose. The most  
37

38 304 common loss of quality was due to the use of accelerometers which assess low movement (e.g.  
39

40 305 ActiGraph) as a criterion measure, as this does not measure the primary aspect of the definition of  
41

42 306 SB (i.e. posture). Another issue which lowered quality was the manipulation of the criterion measure  
43

44 307 without clear justification. For example, some studies manipulated the count threshold (used to  
45

46 308 define SB) or included only SB bouts longer than a particular duration without justification or solid  
47

48 309 rationale.  
49  
50

51 310  
52  
53

54  
55 311 Tools and measurement characteristics  
56  
57  
58  
59  
60

1  
2  
3 312 Table 3 summarises the results reported by these studies, arranged per measurement tool and  
4  
5 313 mapped against the relevant taxon. Very few of the existing tools to measure SB using self-report  
6  
7 314 have actually been investigated for these measurement characteristics. Accuracy has been reported  
8  
9 315 for ten out of the 37 tools identified in the inventory (IPAQ-L 17d, IPAQ-S 17d, GPAQ, MOST, CHAMPS,  
10  
11 316 LASA, PAST, PAST-U, STAQ, SIT-Q-7d). The most tested tools were the IPAQ in its long form, last  
12  
13 317 seven days [16, 51-55] and short form, last seven days.[56-60]. The SIT-Q-7d was tested in three  
14  
15 318 studies [48, 63-64], and the CHAMPS was investigated in two studies [15, 62]. Information for other  
16  
17 319 tools; GPAQ [61], LASA [33], MOST [35], PAST [42], PAST-U [43], and STAQ [50], come from single  
18  
19 320 studies. Reports of sensitivity to change are only available for two tools; MOST [35] and PAST [42].  
20  
21  
22  
23  
24  
25  
26  
27  
28

29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

322 Taxa tested

323 The literature provides measurement characteristics information for eight distinct full taxa:

324 1.1.1/2.2/3.1/4.3 with six studies on IPAQ-L uw;

325 1.1.1/2.2/3.1/4.1 with five studies on IPAQ-S uw;

326 1.1.1/2.4/3.1/4.5 with one study on GPAQ;

327 1.2.2.1/2.1/3.1/4.5 with one study on PAST and one study on PAST-U;

328 1.2.2.1/2.2/3.1/4.3 with three studies on SIT-Q-7d;

329 1.2.2.1/2.2/3.2/4.5 with one study on MOST;

330 1.2.2.1/2.3/3.2/4.5 with two studies on CHAMPS; and

331 1.2.2.1/2.4/3.1/4.3 with one study on LASA and one study on STAQ.

332

1  
2  
3 333 For the assessment type, there is information for direct measures via single item (1.1.1, twelve  
4  
5 334 studies) and for composite sums of behaviours (1.2.2.1, ten studies). However, there is no  
6  
7 335 information for direct proxy measures (1.1.2). For recall period, there is information on all four  
8  
9 336 possible categories (2.1 previous day, two studies; 2.2 previous week, sixteen studies; 2.3 longer,  
10  
11 337 two studies; and 2.4 unanchored, two studies). The unanchored recall period (2.4), used by 40% of  
12  
13 338 the tools in the inventory, is particularly under-represented with only two studies in the validation  
14  
15 339 review. For temporal scale there is mostly information for assessment at day scale (3.1, twenty  
16  
17 340 studies) and only three studies for the temporal scale of a week (3.2). This is broadly representative  
18  
19 341 of usage by tools in the inventory. For assessment period there is information for weekdays only  
20  
21 342 (4.1, five studies) or both weekdays and weekend days (4.3, eleven studies) and for tools with the  
22  
23 343 assessment period not defined (4.5, six studies). The assessment period not defined taxon (4.5),  
24  
25 344 used by over half the tools in the inventory, is under-represented by these validation studies.  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

#### 346 Accuracy

347 Information for taxon 1.1.1/2.2/3.1/4.3 (IPAQ-L-I7d) is not equivocal. The majority of studies  
348 reported a large underestimation of total SB time ranging from 1.1 hours in adults [54] to 6 hours in  
349 older adults [51]. One study reported that tools in this taxon overestimate total SB time by 2.2 hours  
350 in adults [53]. It is clear that the systematic error on estimates of total SB time using tools from this  
351 taxon is likely to be very large (several hours/day). The random error is also likely to be very large as  
352 the Limits of Agreement reported were consistently very large. Information for taxon  
353 1.1.1/2.2/3.1/4.1 (IPAQ-S-I7d) is a little more consistent for adults. Tools in this taxon seem to  
354 underestimate total SB time by 1.5 to 3 hours in adults. However, in older adults this was less clear  
355 with reports of underestimation by 4.4 hours [59] and overestimation by 3.5 hours [56]. In both  
356 populations the error and Limits of Agreement were large, but not as large as for the previous taxon.

1  
2  
3 357  
4  
5  
6 358

7 Although not entirely consistent, tools reporting information from a single item as a direct measure  
8 359 of sitting (taxon 1.1.1) tended to underestimate sitting, with underestimation ranging from -0.5 [60]  
9  
10 360 to -6.0 [51] hours per day. Within those tools, the IPAQ-S-I7d (reporting only for week days in the  
11  
12 361 past week, taxa 2.2 & 4.1), tended to have better agreement than the IPAQ-L-I7d (reporting for both  
13  
14 362 week and weekend days in the past week, taxa 2.2 & 4.3), and the GPAQ (reporting over a longer  
15  
16 363 recall period with the assessment period not defined, taxa 2.4 & 4.5). Tools reporting on a sum of  
17  
18 364 behaviours (taxon 1.2.2.1), were more likely to overestimate sitting than for the single item direct  
19  
20 365 measure (taxon 1.1.1). Tools which reported on a sum of behaviours over the past day or past week  
21  
22 366 (taxa 1.2.2.1 & 2.1 or 2.2), tended to have the closest agreement with objective criterion measures,  
23  
24 367 with most studies reporting agreement between -1.0 and +2.3 hours per day. Tools which reported  
25  
26 368 sum of behaviours over a longer (taxon 2.3) or unanchored (taxon 2.4) recall period or which had a  
27  
28 369 temporal unit of a week (taxon 3.2) reported larger underestimates (-2.1 to -6.8 hours/day). In  
29  
30 370 particular the CHAMPS tool, reporting both for a recall period of a year (taxon 2.3) with a temporal  
31  
32 371 unit of a week (taxon 3.2), had the largest differences for any tool. However, there were only a few  
33  
34 372 studies reporting on these aspects, and such conclusions are necessarily tentative. Regardless of  
35  
36 373 level of agreement, limits of agreement were large for all tools.

37  
38  
39  
40  
41 374  
42  
43

44 375 Sensitivity to change  
45  
46

47 376 There is almost no information about sensitivity to change. The two studies that assessed sensitivity  
48  
49 377 to change [35, 42] provided little tangible information. The results were either inconclusive [42], or  
50  
51 378 reported the Guyatt index against a criterion measure which does not measure sitting [35]. While  
52  
53 379 the latter provided some indication that the tools' sensitivity to change was similar to that of an  
54  
55 380 objective measure of low movement it does not give a clear indication as to whether it is sensitive to  
56  
57  
58  
59  
60

1  
2  
3 381 a change in total SB time. Neither of these studies reported the minimal detectable change [65], a  
4  
5 382 metric which provides an easily interpretable value of the capacity of a tool to detect a change.  
6  
7

8 383  
9

## 10 384 **DISCUSSION**

11  
12  
13  
14 385 A taxonomy (TASST) for the systematic description and comparison of self-reported measures of SB  
15  
16 386 has been established. TASST provides a rigorous framework for informed choice, development and  
17  
18 387 evaluation of self-report tools. This framework has been used to review the measurement  
19  
20 388 characteristics of existing tools in order to identify the optimum tool for population surveillance. The  
21  
22 389 available evidence about measurement characteristics essential for population surveillance, namely  
23  
24 390 accuracy and responsiveness to change, was insufficient to ascertain which tool currently used in  
25  
26 391 practice is best. Accuracy was poor for all existing tools, with both under and over estimation of  
27  
28 392 total time spent in SB and large limits of agreement. In addition, there is a complete lack of evidence  
29  
30 393 about their sensitivity to change. Mapping available evidence onto the TASST framework has  
31  
32 394 enabled informed recommendations to be made about the promising features for a surveillance  
33  
34 395 tool, and identification of the aspects on which future research and development of SB surveillance  
35  
36 396 tools should focus.  
37

38  
39  
40  
41 397  
42

43  
44 398 The use of a coherent and robust taxonomy (TASST) to systematically evaluate and compare the  
45  
46 399 characteristics of measurement tools is the main strength of this study. However, in terms of  
47  
48 400 accuracy and sensitivity to change, the current published evidence does not cover the entire  
49  
50 401 taxonomy. Consequently, at present, only tentative recommendations can be provided. The  
51  
52 402 taxonomy can be used, however, to identify gaps in current research and provide focussed guidance  
53  
54 403 for future research and development. During the development of TASST, self-report tools which  
55  
56 404 aimed to measure SB in specific populations (e.g. children, those with arthritis) or specialised  
57  
58  
59  
60

1  
2  
3 405 contexts (e.g. workplace) were not considered. However, TASST is a generic framework, so tools  
4  
5 406 specific to these populations may already be fully described by the taxonomy. For example, a  
6  
7 407 question asking about time spent sitting at school which is specific to children, would be covered  
8  
9 408 under the sub-division of the day assessment period (taxon 4.4). Another consequence of the  
10  
11 409 exclusion criteria is that evidence on accuracy and sensitivity to change of tools specific to these  
12  
13 410 populations was not mapped on the taxonomy. Therefore, the conclusions drawn from the  
14  
15 411 measurement characteristics in this study are only valid for adults and older adults. Future research  
16  
17 412 should be conducted using the TASST taxonomy to map existing self-report tools covering those  
18  
19 413 populations and contexts currently excluded from taxonomy development (such as children, schools  
20  
21 414 or the workplace) to identify areas for development. In addition, this study has the general  
22  
23 415 limitations common to most systematic reviews, i.e. included articles were restricted to those  
24  
25 416 written in English, articles and tools published after the date of search were not included, and any  
26  
27 417 relevant articles not identified during the search will have been excluded.  
28  
29  
30  
31  
32 418

33  
34  
35 419 The current study is the first to clearly define and focus on the measurement characteristics required  
36  
37 420 for population surveillance (accuracy and sensitivity to change). There is only one other systematic  
38  
39 421 review reporting on the measurement characteristics of self-report tools to measure SB [13], which  
40  
41 422 concentrated on validity (assessed through rank correlation) and reliability, which are the  
42  
43 423 measurement characteristics relevant to establishing associations between SB and health. In  
44  
45 424 agreement with the previous review, we found that the major flaw of most validation studies was  
46  
47 425 the use of an inadequate criterion measure. The choice of criterion measure depends on the  
48  
49 426 purpose of the tool. While direct observation should be considered the gold standard, if the purpose  
50  
51 427 is to assess total sedentary time, then accurate postural sensors should be adequate (e.g. activPAL).  
52  
53 428 In this review, only five out of 22 studies used an adequate criterion measure. Instead, many studies  
54  
55 429 used an accelerometer which measures low levels of movement at the hip (e.g. ActiGraph) as a  
56  
57  
58  
59  
60

1  
2  
3 430 criterion measure, but such tools do not measure SB directly and can misclassify standing as sitting  
4  
5 431 [12]. Although it is possible that criterion measure may have provided a confounding effect on  
6  
7 432 agreement (e.g. tools assessing previous day recall period (taxon 2.1, PAST, PAST-U) were only  
8  
9 433 assessed against the activPAL), no clear trend towards better or worse agreement with a particular  
10  
11 434 type of criterion measure or ActiGraph cut-off was apparent.  
12  
13

435

14  
15  
16  
17 436 Despite the incomplete nature of the evidence, TASST enables the identification of desirable  
18  
19 437 characteristics of self-report tools to measure SB when used for population surveillance. Firstly,  
20  
21 438 tools assessing total SB time as a sum of behaviours (taxon 1.2.2.1; CHAMPS, LASA, MOST, PAST,  
22  
23 439 PAST-U, SIT-Q-7d, STAQ) provided better accuracy than single item direct measurement (taxon 1.1.1;  
24  
25 440 IPAQ-L-I7d, IPAQ-S-I7d and GPAQ) tools, especially when comparing tools with equivalent recall  
26  
27 441 periods. However, this will be dependent on the behaviours or domains included within the sum,  
28  
29 442 and whether they are exhaustive, consistent and mutually exclusive. Tools with a non-exhaustive  
30  
31 443 sum will underestimate total time, for example, the Longitudinal Aging Study Amsterdam (LASA),  
32  
33 444 found that a six item sum provided a better correlation with SB across the sample, but that a ten  
34  
35 445 item sum was more accurate [33]. Conversely, tools which contain behaviours which might occur  
36  
37 446 concurrently (such as watching TV and using a tablet computer) may lead to an over-estimate in  
38  
39 447 total SB time [63]. Secondly, tools using a previous day recall period (taxon 2.1, PAST, PAST-U)  
40  
41 448 tended to provide better accuracy than those with longer recall periods (taxa 2.2, 2.3 and 2.4). This  
42  
43 449 corroborates recent research on the validity of computerised survey systems which assess SB using a  
44  
45 450 past-day recall period [17, 66]. However, although tools using previous day recall may more  
46  
47 451 accurate, it is likely that their sensitivity to change will be less good due to the higher underlying  
48  
49 452 variability in daily SB [67].  
50  
51  
52  
53  
54

453



1  
2  
3 454 Most tools currently used for population surveillance of SB systematically underestimate the amount  
4  
5 455 of SB by two to four hours per day. Yet, self-report tools are still the most practical and economical  
6  
7 456 means of population surveillance. Therefore, policy makers and clinicians should be aware that  
8  
9 457 reports of population SB time are likely to be grossly underestimated, and should be cognisant of  
10  
11 458 this fact when making decisions on implementing, developing and evaluating policy and public  
12  
13 459 health interventions. In addition, policy makers and clinicians should be cautious in interpreting any  
14  
15 460 reported difference in population SB time as a real change. The dearth of information about  
16  
17 461 sensitivity to change of these tools means that we do not know the magnitude of change required to  
18  
19 462 be certain that a change is real and not background variation. Moving forward, development of  
20  
21 463 national and international surveillance systems should not be undertaken assuming that a tool is  
22  
23 464 adequate because it has been used previously. Instead, investment should be made in research to  
24  
25 465 evaluate the sensitivity to change and accuracy of tools to measure SB, paying attention to the  
26  
27 466 potential trade-off between these two measurement characteristics. Such research should be  
28  
29 467 carefully planned, to ensure that meaningful comparisons are investigated. The TASST taxonomy  
30  
31 468 should be used as a useful framework to facilitate such a systematic approach.  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

469

1  
2  
3 470 Acknowledgements  
4  
5

6 471 The named authors present the study on behalf of the Seniors USP Team, which comprises: Dawn A  
7  
8 472 Skelton (PI), Sebastien Chastin, Simon Cox, Elaine Coulter, Iva Čukić, Philippa Dall, Ian Deary, Geoff  
9  
10 473 Der, Manon Dontje, Claire Fitzsimons, Catherine Gale, Jason Gill, Malcolm Granat, Cindy Gray,  
11  
12 474 Carolyn Greig, Elaine Hindle, Karen Laird, Gillian Mead, Nanette Mutrie, Victoria Palmer, Ratko  
13  
14 475 Radakovic, Naveed Sattar, Richard Shaw, John Starr, Sally Stewart, Sally Wyke. The authors would  
15  
16 476 like to thank Calum Leask (CL) for his help in updating both of the searches in November 2016.  
17  
18  
19

20 477  
21

22 478 Competing Interests:  
23  
24

25 479 All authors had financial support from the Medical Research Council for the submitted work; no  
26  
27 480 financial relationships with any organisations that might have an interest in the submitted work in  
28  
29 481 the previous three years; no other relationships or activities that could appear to have influenced  
30  
31 482 the submitted work.  
32  
33

34 483  
35  
36  
37

38 484 Funding:  
39  
40

41 485 The study was performed in preparation for the Seniors USP: Understanding Sedentary Patterns  
42  
43 486 project, which was supported by the Medical Research Council [MR/K025023/1]  
44  
45

46 487  
47  
48

49 488 Authors Contributions:  
50  
51

52 489 SC and DS conceived and supervised the study. PD and SC designed the study. PD, EC, CF and SC  
53  
54 490 collected the data, and performed the systematic review. PD, CF and SC analysed the data for the  
55  
56  
57  
58  
59  
60

1  
2  
3 491 taxonomy. PD and SC analysed the data from the systematic review. All authors interpreted the  
4  
5 492 data and critically revised the manuscript for important intellectual content.  
6

7 493  
8  
9

10 494 Data Sharing: Data can be obtained from the corresponding author on request.  
11  
12 495

13  
14  
15 496 Figure Legends  
16

17 497  
18  
19

20  
21 498 Figure 1: TAxonomy of Self-reported Sedentary behaviour Tools (TASST)  
22  
23

24 499  
25  
26

27 500 Figure 2: PRISMA diagram of the validation systematic review  
28  
29

30 501  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 502 References  
45  
6 503  
78  
9 504 1. Kohl HW 3rd, Craig CL, Lambert EV, et al. The pandemic of physical inactivity: global action for  
10  
11 505 public health. *Lancet* 2012;380:294–305.  
1213  
14 506 2. Chastin S, Schwartz U, Skelton D. Development of a consensus taxonomy of sedentary behaviors  
15  
16 507 (SIT): report of Delphi round 1. *PLoS One* 2013;8:e82313.  
1718  
19 508 3. Sedentary Behaviour Research Network. Standardized use of the terms “sedentary” and  
20  
21 509 “sedentary behaviours”. *Appl Physiol Nutr Metabol* 2012;37:540-2.  
2223  
24 510 4. Healy GN, Matthews CE, Dunstan DW, Winkler EAH, Owen N. Sedentary time and cardio-  
25  
26 511 metabolic biomarkers in US adults: NHANES 2003-06. *Eur Heart J* 2011;32:590-7.  
2728  
29 512 5. Harvey JA, Chastin SFM, Skelton DA. How sedentary are older people? A systematic review of the  
30  
31 513 amount of sedentary behavior. *J Aging Phys Act* 2015;23:471-87.  
3233  
34 514 6. Proper KI, Singh AS, van Mechelen W, Chinapaw MJM. Sedentary behaviors and health outcomes  
35  
36 515 among adults: a systematic review of prospective studies. *Am J Prev Med* 2011;40:174-82.  
3738  
39 516 7. Thorp AA, Owen N, Neuhaus M, Dunstan DW. Sedentary behaviors and subsequent health  
40  
41 517 outcomes in adults a systematic review of longitudinal studies, 1996-2011. *Am J Prev Med*  
42  
43 518 2011;41:207-15.  
4445  
46 519 8. de Rezende LFM, Rey-López JP, Matsudo VKR, do Carmo Luiz O. Sedentary behavior and health  
47  
48 520 outcomes among older adults: a systematic review. *BMC Public Health* 2014;14:333.  
4950  
51 521 9. Dogra S, Stathokostas L. Sedentary behavior and physical activity are independent predictors of  
52  
53 522 successful aging in middle-aged and older adults. *J Aging Res* 2012;2012:190654.  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 523 10. Department of Health. Start Active, Stay Active: A report on physical activity from the four home  
4  
5 524 countries' Chief Medical Officers. 2011.  
6  
7  
8 525 11. Guyatt G, Walter S, Norman G. Measuring change over time. assessing the usefulness of  
9  
10 526 evaluative instruments. *J Chron Dis* 1987;40:171-8.  
11  
12  
13 527 12. Kozey-Keadle S, Libertine A, Lyden K, Staudenmayer J, Freedson P. Validation of wearable  
14  
15 528 monitors for assessing sedentary behavior. *Med Sci Sports Exerc* 2011;43:1561-7.  
16  
17  
18 529 13. Healy GN, Clark BK, Winkler EAH, Gardiner PA, Brown WJ, Matthews CE. Measurement of adults'  
19  
20 530 sedentary time in population-based studies. *Am J Prev Med* 2011;41:216-27.  
21  
22  
23 531 14. Bauman A, Ainsworth BE, Sallis JF, et al. The descriptive epidemiology of sitting: a 20 country  
24  
25 532 comparison using the International Physical Activity Questionnaire (IPAQ). *Am J Prev Med*  
26  
27 533 2011;41:228-35.  
28  
29  
30 534 15. Hekler EB, Buman MP, Haskell WL, et al. Reliability and validity of CHAMPS self-reported  
31  
32 535 sedentary-vigorous intensity physical activity in older adults. *J Phys Act Health* 2012;9:225-36.  
33  
34  
35 536 16. Chastin SFM, Culhane B, Dall PM. Questionnaire sitting time (IPAQ) compared to objective  
36  
37 537 measurement via inclinometry (activPAL). *Physiol Meas* 2014;35:2319-28.  
38  
39  
40 538 17. Kozey-Keadle S, Lyden K, Hickey A, et al. Validation of a previous day recall for measuring the  
41  
42 539 location and purpose of active and sedentary behaviours compared to direct observation. *Int J*  
43  
44 540 *Behav Nutr Phys Act* 2014;11:12.  
45  
46  
47  
48 541 18. Kmet LM, Lee RC, Cook LS. Standard quality assessment criteria for evaluating primary research  
49  
50 542 papers from a variety of fields. *Edmonton: Alberta Heritage Foundation for Medical Research*  
51  
52 543 2004;1-22.  
53  
54  
55 544 19. Banks E, Jorm L, Rogers K, Clements M, Bauman A. Screen-time, obesity, ageing and disability:  
56  
57 545 findings from 91 266 participants in the 45 and Up Study. *Public Health Nutr* 2011;14:34-43.  
58  
59  
60

- 1  
2  
3 546 20 Patel AV, Bernstein L, Deka A, et al. Leisure time spent sitting in relation to total mortality in a  
4  
5 547 prospective cohort of US adults. *Am J Epidemiol* 2010;172:419-429.  
6  
7  
8 548 21. Salmon J, Owen N, Crawford D, Bauman A, Sallis JF. Physical activity and sedentary behavior: a  
9  
10 549 population-based study of barriers, enjoyment, and preference. *Health Psychol* 2003;22:178-88.  
11  
12  
13 550 22. Gardiner PA, Healy GN, Eakin EG, et al. Associations between television viewing time and overall  
14  
15 551 sitting time with the metabolic syndrome in older men and women: The Australian Diabetes Obesity  
16  
17 552 and Lifestyle study. *JAGS* 2011;59:788-96.  
18  
19  
20 553 23. Shields M, Tremblay MS. Sedentary behaviour and obesity. *Health Rep.* 2008;19:19-30.  
21  
22  
23 554 24. Katzmarzyk PT, Church TS, Craig Cl, Bouchard C. Sitting time and mortality from all causes,  
24  
25 555 cardiovascular disease, and cancer. *Med Sci Sports Exerc* 2009;41:998-1005.  
26  
27  
28 556 25. Hamer M, Stamatakis E. Screen-based sedentary behaviour, physical activity, and muscle  
29  
30 557 strength in the English Longitudinal Study of Ageing. *PLoS One* 2013;8:e66222.  
31  
32  
33 558 26. Wareham NJ, Jakes RW, Rennie KL, Mitchell J, Hennings S, Day NE. Validity and repeatability of  
34  
35 559 the EPIC-Norfolk physical activity questionnaire. *Int J Epidemiol* 2002;31:186-74.  
36  
37  
38 560 27. Parsons TJ, Thomas C, Power C. Estimated activity patterns in British 45 year olds: cross-  
39  
40 561 sectional findings from the 1958 British birth cohort. *Eur J Clin Nutr* 2009;63:978-85.  
41  
42  
43 562 28. Armstrong T, Bull F. Development of the World Health Organization Global Physical Activity  
44  
45 563 Questionnaire (GPAQ). *J Public Health* 2006;14:66-70.  
46  
47  
48 564 29. Stamatakis E, Davis M, Stathi A, Hamer M. Associations between multiple indicators of  
49  
50 565 objectively-measured and self-reported sedentary behaviour and cardiometabolic risk in older  
51  
52 566 adults. *Prev Med* 2012;54:82-7.  
53  
54  
55  
56  
57  
58  
59  
60

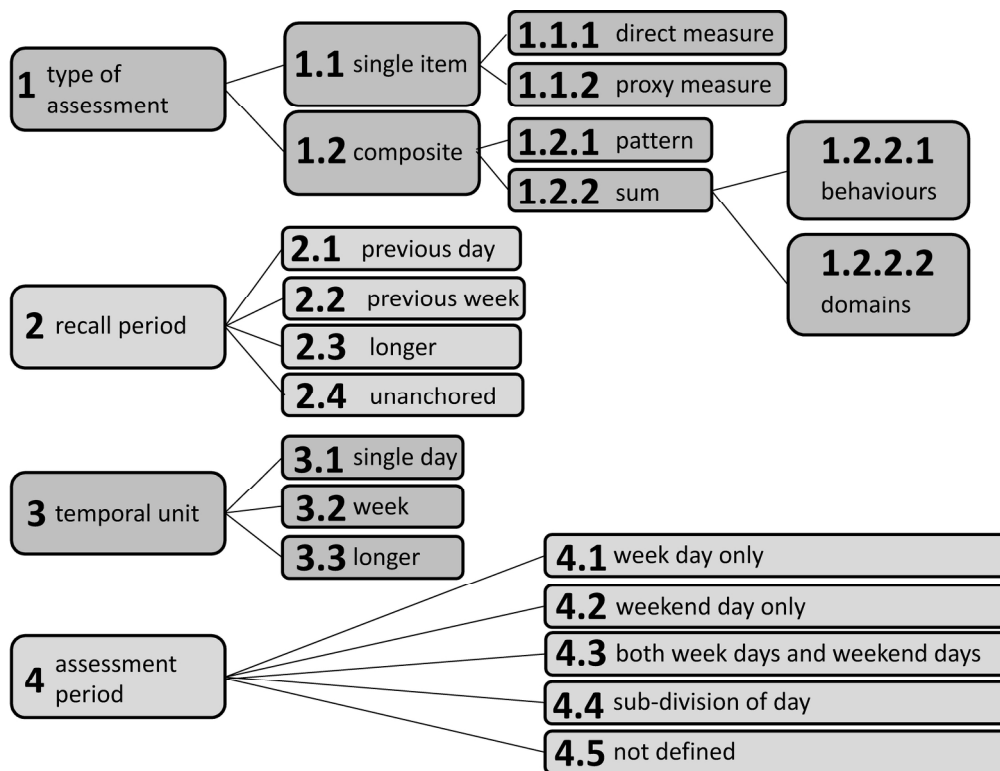
- 1  
2  
3 567 30. Chau JY, Grunseit A, Midthjell K, et al. Sedentary behaviour and risk of mortality from all-causes  
4  
5 568 and cardiometabolic diseases in adults: evidence from the HUNT3 population cohort. *Br J Sports*  
6  
7 569 *Med* 2015;49:737-42.
- 8  
9  
10 570 31. Craig CL, Marshall AL, Sjöström M, et al. International Physical Activity Questionnaire: 12-country  
11  
12 571 reliability and validity. *Med Sci Sports Exerc* 2003;35:1381-95.
- 13  
14  
15 572 32. Proper KI, Cerin E, Brown WJ, Owen N. Sitting time and socio-economic differences in overweight  
16  
17 573 and obesity. *Int J Obes* 2007;31:169-76.
- 18  
19  
20 574 33. Visser M, Koster A. Development of a questionnaire to assess sedentary time in older persons – a  
21  
22 575 comparative study using accelerometry. *BMC Geriatr* 2013;13:80.
- 23  
24  
25 576 34. Burazeri G, Goda A, Kark JD. Television viewing, leisure-time exercise and acute coronary  
26  
27 577 syndrome in transitional Albania. *Prev Med* 2008;47:112-5.
- 28  
29  
30 578 35. Gardiner PA, Clark BK, Healy GN, Eakin EG, Winkler EAH, Owen N. Measuring older adults'  
31  
32 579 sedentary time: reliability, validity, and responsiveness. *Med Sci Sports Exerc* 2011;43: 2127-33.
- 33  
34  
35 580 36. Ford ES. Combined television viewing and computer use and mortality from all-causes and  
36  
37 581 diseases of the circulatory system among adults in the United States. *BMC Public Health* 2012;12:70
- 38  
39  
40 582 37. Evenson KR, McGinn AP. Test-retest reliability of adult surveillance measures for physical activity  
41  
42 583 and inactivity. *Am J Prev Med* 2005;28:470-8.
- 43  
44  
45 584 38. Wolf AM, Hunter DJ, Colditz GA, et al. Reproducibility and validity of a self-administered physical  
46  
47 585 activity questionnaire. *Int J Epidemiol* 1994;23:991-99.
- 48  
49  
50 586 39. Matthews CE, George SM, Moore SC, et al. Amount of time spent in sedentary behaviors and  
51  
52 587 cause-specific mortality in US adults. *Am J Clin Nutr* 2012;ajcn-019620.
- 53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 588 40. Salmon J, Bauman A, Crawford D, Timperio A, Owen N. The association between television  
4  
5 589 viewing and overweight among Australian adults participating in varying levels of leisure-time  
6  
7 590 physical activity. *Int J Obes Relat Metab Disord* 2000;24:600-6.  
8  
9  
10 591 41. Washburn RA, McAuley E, Katula J, Mihalko SL, Boileau RA. The Physical Activity Scale for the  
11  
12 592 Elderly (PASE): evidence for validity. *J Clin Epidemiol* 1999;52:643-51.  
13  
14  
15 593 42. Clark BK, Winkler E, Healy GN, et al. Adults' past-day recall of sedentary time: reliability, validity  
16  
17 594 and responsiveness. *Med Sci Sports Exerc* 2013;45:1198-207.  
18  
19  
20 595 43. Clark BK, Pavey TG, Lim RF, Gomersall SR, Brown WJ. Past-day recall of sedentary time: validity  
21  
22 596 of a self-reported measure of sedentary time in a university population. *J Sci Med Sport* 2016;19:237-  
23  
24 597 41.  
25  
26  
27 598 44. Balboa-Castillo T, León-Muñoz LM, Graciani A, Rodríguez-Artalejo F, Guallar-Castillón P.  
28  
29 599 Longitudinal association of physical activity and sedentary behavior during leisure time with health-  
30  
31 600 related quality of life in community-dwelling older adults. *Health Qual Life Outcomes* 2011;9:41-47.  
32  
33  
34 601 45. Rosenberg DE, Norman GK, Wagner N, Patrick K, Calfas KJ, Sallis JF. Reliability and validity of the  
35  
36 602 Sedentary Behaviour Questionnaire (SBQ) for adults. *J Phys Act Health* 2010;7:697-705.  
37  
38  
39 603 46. Stamatakis E, Hamer M, Dunstan DW. Screen-based entertainment time, all-cause mortality, and  
40  
41 604 cardiovascular events. *J Am Coll Cardiol* 2011;57:292-9.  
42  
43  
44 605 47. Lynch BM, Friedenreich CM, KHandwala F, Liu A, Nicholas J, Csizmadia I. Development and testing  
45  
46 606 of a past year measure of sedentary behaviour: the SIT-Q. *BMC Public Health*; 2014;14:899.  
47  
48  
49 607 48. Wijndaele K, de Bourdeaudhuij I, Godino JG, et al. Reliability and validity of a domain-specific  
50  
51 608 last 7-d sedentary time questionnaire. *Med Sci Sports Exerc* 2014;46:1248-60.  
52  
53  
54  
55  
56  
57  
58  
59  
60



- 1  
2  
3 609 49. Csizmadi I, Neilson HK, Kopciuk KA, et al. The Sedentary Time and Activity reporting  
4  
5 610 Questionnaire (STAR-Q): reliability and validity against doubly labelled water and 7-day activity  
6  
7 611 diaries. *Am J Epidemiol* 2014;80:424-35.  
8  
9  
10 612 50. Mensah K, Maire A, Oppert J-M, et al. Assessment of sedentary behaviours and transport-  
11  
12 613 related activities by questionnaire: a validation study. *BMC Public Health* 2016;16:753.  
13  
14  
15 614 51. Dahl-Pederson KI, Hansen AW, Bjerregaard P, Jørgensen ME, Brage S. Validity of the  
16  
17 615 international physical activity questionnaire in the arctic. *Med Sci Sports Exerc* 2013;45:728-36.  
18  
19  
20 616 52. van Dyck D, Cardon G, Deforche B, de Bourdeaudhuij I. IPAQ interview version: convergent  
21  
22 617 validity with accelerometers and comparison of physical activity and sedentary time levels with the  
23  
24 618 self-administered version. *J Sports Med Phys Fitness* 2014;55:776-86.  
25  
26  
27  
28 619 53. Hagstromer M, Ainsworth BE, Sjostrom M. Comparison of a subjective and an objective measure  
29  
30 620 of physical activity in a population sample. *J Phys Act Health* 2010;7:541-50.  
31  
32  
33 621 54. Celis-Morales CA, Perez-Bravo F, Ibanez L, Salas C, Bailey MES, Gill JMR. Objective vs. self-  
34  
35 622 reported physical activity and sedentary time: effects of measurement method on relationships with  
36  
37 623 risk biomarkers. *PLoS ONE* 2012;7:e36345.  
38  
39  
40 624 55. Wanner M, Probst-Hensch N, Kriemler S, Meier F, Autenrieth C, Martin BW. Validation of the  
41  
42 625 long international physical activity questionnaire: influence of age and language region. *Prev Med*  
43  
44 626 *Rep* 2016;3:250-6.  
45  
46  
47 627 56. Dyrstad SM, Hansen BH, Holme IM, Anderssen SA. Comparison of self-reported versus  
48  
49 628 accelerometer-measured physical activity. *Med Sci Sports Exerc* 2014;46:99-106.  
50  
51  
52 629 57. Oyeyemi AL, Umar M, Oguiche F, Aliyu SU, Oyeyemi AY. Accelerometer-determined physical  
53  
54 630 activity and its comparison with the International Physical Activity Questionnaire in a sample of  
55  
56 631 Nigerian adults. *PLoS ONE* 2014;9:e87233.  
57  
58  
59  
60

- 1  
2  
3 632 58. Hurtig-Wennlöf A, Hagströmer M, Olsson L. The International Physical Activity Questionnaire  
4  
5 633 modified for the elderly: aspects of validity and feasibility. *Public Health Nutr* 2010;13:1847-54.  
6  
7  
8 634 59. Grimm EK, Swartz AM, Hart T, Miller NE, Strath SJ. Comparison of the IPAQ-short form and  
9  
10 635 accelerometry predictions of physical activity in older adults. *J Aging Phys Act* 2012;20:64-79.  
11  
12  
13 636 60. Curry WB, Thompson JL. Comparability of Accelerometer- and IPAQ-derived physical activity and  
14  
15 637 sedentary time in south Asian women: a cross-sectional study. *Eur J Sport Sci* 2015;15:655-62.  
16  
17  
18 638 61. Alkahtani SA. Convergent validity: agreement between accelerometry and the Global Physical  
19  
20 639 Activity Questionnaire in college-age Saudi men. *BMC Res Notes* 2016;9:436.  
21  
22  
23 640 62. Gennuso KP, Matthews CE, Colbert LH. Reliability and validity of two self-report measures to  
24  
25 641 assess sedentary behavior in older adults. *J Phys Act Health* 2015;12:727-32.  
26  
27  
28 642 63. Busschaert C, de Bourdeaudhuij I, van Holle V, Chastin SFM, Cardon G, de Cocker K. Reliability  
29  
30 643 and validity of three questionnaires measuring context-specific sedentary behaviour and associated  
31  
32 644 correlates in adolescents, adults, and older adults. *Int J Behav Nutr Phys Act* 2015;12:117.  
33  
34  
35 645 64. van Cauwenberg J, van Holle V, de Bourdeaudhuij I, Owen N, Deforche B. Older adults' reporting  
36  
37 646 of specific sedentary behaviours: validity and reliability. *BMC Public Health* 2014;14:734.  
38  
39  
40 647 65. Beaton DE, Bombardier C, Katz JN, Wright JG. A taxonomy for responsiveness. *J Clin Epidemiol*  
41  
42 648 2001;54:1204-71.  
43  
44  
45 649 66. Gomersall SR, Pavey T, Clark BK, Jasman A, Brown WJ. Validity of a self-report recall tool for  
46  
47 650 estimating sedentary behaviour in adults. *J Phys Act Health* 2015;12:1485-91.  
48  
49  
50 651 67. Hart TL, Swartz AM, Cashin SE, Strath SJ. How many days of monitoring predict physical activity  
51  
52 652 and sedentary behaviour in older adults? *Int J Behav Nutr Phys Act* 2011;8:62.  
53  
54  
55  
56  
57  
58  
59  
60

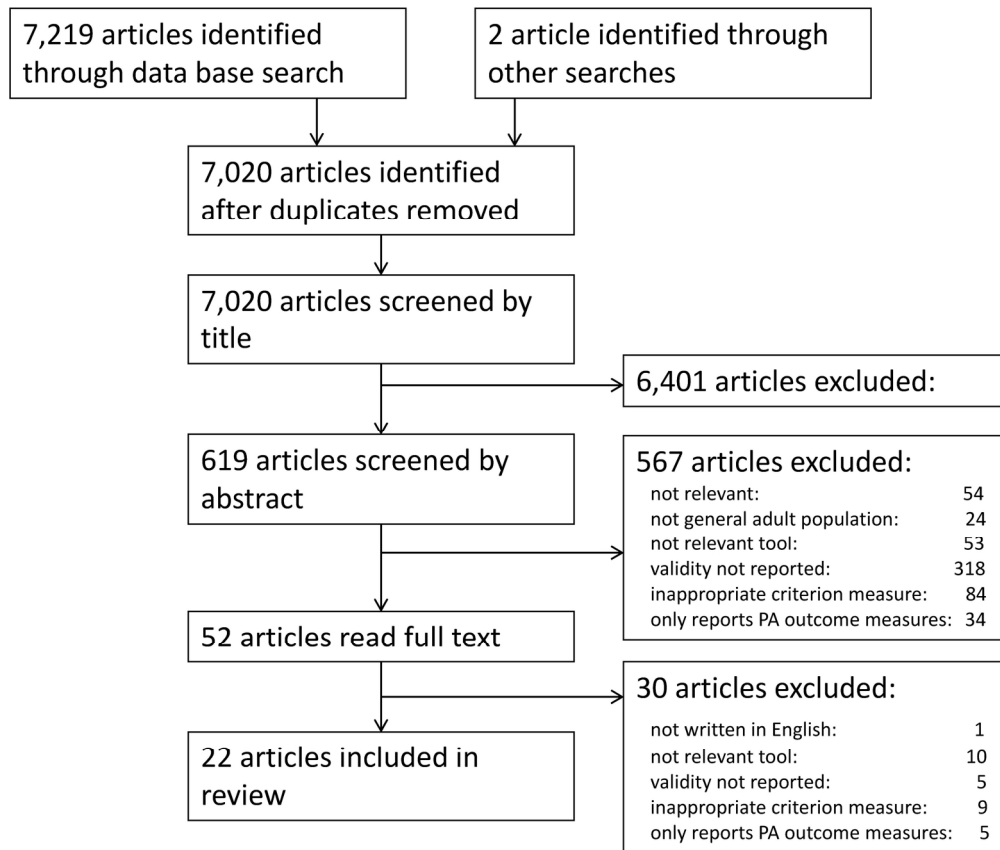


Taxonomy of Self-reported Sedentary behaviour Tools (TASST)

Figure 1

186x142mm (300 x 300 DPI)

ew only



PRISMA diagram of the validation systematic review

Figure 2

171x144mm (300 x 300 DPI)

For peer review only



# PRISMA 2009 Checklist

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

Section/topic	#	Checklist item	Reported on page #
<b>TITLE</b>			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
<b>ABSTRACT</b>			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
<b>INTRODUCTION</b>			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	3
<b>METHODS</b>			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	4
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	4 & 5
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	4 & 5
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	5
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	5
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	5
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	5
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	Not reporting on trials, so not done



# PRISMA 2009 Checklist

Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	5
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., $I^2$ ) for each meta-analysis.	Not done

Page 1 of 2

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	9
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	Not done
<b>RESULTS</b>			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	8, figure 2
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Table 3
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	8
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Table 3, pages 9 & 10 (not a review of interventions)
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	Not enough data for meta-analysis
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	9
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	Not done
<b>DISCUSSION</b>			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	12
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	11 & 12



# PRISMA 2009 Checklist

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	12
<b>FUNDING</b>			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	14

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: [www.prisma-statement.org](http://www.prisma-statement.org).

Page 2 of 2