

Prediction of objectively measured physical activity and sedentariness among blue-collar workers using survey questionnaires

Gupta, Nidhi; Heiden, Marina; Mathiassen, Svend Erik; Holtermann, Andreas

Published in:
Scandinavian Journal of Work, Environment & Health

DOI:
10.5271/sjweh.3561

Publication date:
2016

Document version:
Final published version

Citation for pulished version (APA):
Gupta, N., Heiden, M., Mathiassen, S. E., & Holtermann, A. (2016). Prediction of objectively measured physical activity and sedentariness among blue-collar workers using survey questionnaires. *Scandinavian Journal of Work, Environment & Health*, 42(3), 237-245. <https://doi.org/10.5271/sjweh.3561>

Go to publication entry in University of Southern Denmark's Research Portal

Terms of use

This work is brought to you by the University of Southern Denmark.
Unless otherwise specified it has been shared according to the terms for self-archiving.
If no other license is stated, these terms apply:

- You may download this work for personal use only.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying this open access version

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim.
Please direct all enquiries to puresupport@bib.sdu.dk



Original article

Scand J Work Environ Health [2016;42\(3\):237-245](#)

doi:10.5271/sjweh.3561

Prediction of objectively measured physical activity and sedentariness among blue-collar workers using survey questionnaires

by [Gupta N](#), [Heiden M](#), [Mathiassen SE](#), [Holtermann A](#)

Refers to the following text of the Journal: [2007;33\(6\):401-480](#)

Key terms: [accelerometry](#); [actigraph](#); [blue-collar worker](#); [objective measure](#); [physical activity](#); [prediction](#); [prediction model](#); [questionnaire](#); [sedentariness](#); [sedentary time](#)

This article in PubMed: www.ncbi.nlm.nih.gov/pubmed/27067991

Additional material

Please note that there is additional material available belonging to this article on the [Scandinavian Journal of Work, Environment & Health -website](#).

Prediction of objectively measured physical activity and sedentariness among blue-collar workers using survey questionnaires

by Nidhi Gupta, PhD,¹ Marina Heiden, PhD,² Svend Erik Mathiassen, PhD,² Andreas Holtermann, PhD^{1,3}

Gupta N, Heiden M, Mathiassen SE, Holtermann A. Prediction of objectively measured physical activity and sedentariness among blue-collar workers using survey questionnaires. *Scand J Work Environ Health*. 2016;42(3):237–245.

Objectives We aimed at developing and evaluating statistical models predicting objectively measured occupational time spent sedentary or in physical activity from self-reported information available in large epidemiological studies and surveys.

Methods Two-hundred-and-fourteen blue-collar workers responded to a questionnaire containing information about personal and work related variables, available in most large epidemiological studies and surveys. Workers also wore accelerometers for 1–4 days measuring time spent sedentary and in physical activity, defined as non-sedentary time. Least-squares linear regression models were developed, predicting objectively measured exposures from selected predictors in the questionnaire.

Results A full prediction model based on age, gender, body mass index, job group, self-reported occupational physical activity (OPA), and self-reported occupational sedentary time (OST) explained 63% (R^2 adjusted) of the variance of both objectively measured time spent sedentary and in physical activity since these two exposures were complementary. Single-predictor models based only on self-reported information about either OPA or OST explained 21% and 38%, respectively, of the variance of the objectively measured exposures. Internal validation using bootstrapping suggested that the full and single-predictor models would show almost the same performance in new datasets as in that used for modelling.

Conclusions Both full and single-predictor models based on self-reported information typically available in most large epidemiological studies and surveys were able to predict objectively measured occupational time spent sedentary or in physical activity, with explained variances ranging from 21–63%.

Key terms accelerometry; actigraph; blue-collar worker; objective measure; prediction model; sedentary time.

Even in modern information societies, a considerable proportion of the working population is exposed to physical activity at work (1, 2). In a national survey in 2012, 39% of the Danish workforce reported to have a job where $\geq 75\%$ of the time required some physical activity, such as standing and walking (3). More self-reported time spent in physical activity during work has been associated with increased risk of long-term sickness absence (2, 4), premature drop-out from the workforce (5), and cardiovascular and all-cause mortality (1, 6, 7). On the other hand, other workers spend a large proportion of the time at work being sedentary (8–10), which has been suggested to be associated with increased all-cause mortality (11), musculoskeletal pain (12), and obesity (13).

Occupational time spent sedentary and in physical activity have so far mainly been determined using questionnaires that are feasible to administer in a large population, such as in national surveys (14–16). However, questionnaires have been criticized for giving biased and imprecise results compared to objective measurements (17). Systematic and random measurement errors may lead to misleading results, both when documenting time spent sedentary and in physical activity and when determining associations with relevant outcomes such as health and well-being. As an alternative, objective measurements using accelerometers offer accurate information of time spent sedentary and in physical activity (18, 19). Thus, accelerometer recordings have been used as the gold standard for validating questionnaire-based data on time spent sedentary

¹ National Research Centre for the Working Environment, Copenhagen, Denmark.

² Centre for Musculoskeletal Research, Department of Occupational and Public Health Sciences, University of Gävle, Gävle, Sweden.

³ Institute of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark.

and in physical activity (20, 21). However, accelerometers demanding more resources to use than questionnaires (22), disqualifying them from most large-scale studies.

An attractive compromise would be to predict objectively measured occupational time spent sedentary and in physical activity from self-reported information that would generally be available in most large epidemiological studies and surveys. Explicit prediction models have been proposed before to predict time spent sedentary and in physical activity (23–25), but these studies have not developed models for exposures at work, which may show associations with self-reported predictors other than leisure time exposures. A few previous studies have, indeed, developed prediction models for time spent sedentary and in physical activity specifically at work (26–30). However, they have mainly focused on predicting answers to some self-reported variables by another type of self-reported information. This approach increases the risk of correlated error or common-method bias (31).

Another limitation of previous prediction models addressing time spent sedentary and in physical activity at work is that the predictors included in the models, such as cognitive (32) and psychosocial variables – including social norms, self-efficacy and advantages of sitting less (26) – are not normally available in large epidemiological studies and surveys. Developing models based on predictors that typically appear in large epidemiological studies and surveys would increase the utility of the models in the context of, for instance, public health surveys and cohort studies of occupational health.

As a general endeavor in exposure modelling, examination of simple models based on few predictors is of interest, since parsimonious models may be easier to use and more stable than models based on many predictors. In the present context, this would call for assessments of the performance of models based only on selected questionnaire variables that can be expected to be particularly predictive of sedentary behavior and physical activity at work. Thus, this study aimed at developing and evaluating statistical models predicting objectively measured time spent sedentary and in physical activity at work from self-reported variables which would generally be available in large epidemiological studies and surveys. A secondary aim was to examine the extent to which single-predictor models based on questions regarding occupational sedentary time (OST) or occupational physical activity (OPA) can predict the result of objective measurements of corresponding variables.

Methods

Study design and population

Recruitment flow of the study population is shown

in Appendix A (www.sjweh.fi/data_repository.php). Workers were recruited from seven blue-collar occupations (manufacturing, assembling, construction, cleaning, garbage collection, mobile plant operation, and health services) in the cross-sectional New Method for Objective Measurements of Physical Activity in Daily Living (NOMAD) study (33) to obtain a wide range of exposures while maintaining homogeneity among workers with respect to socioeconomic status.

The Ethics Committee for the Capital Region in Denmark approved the study (journal number H-2-2011-047), which was conducted in accordance with the Helsinki declaration.

Procedure

At each workplace, data were collected continuously during a four-day period, with research staff being present at the workplace on days one and four (33, 34). On day one, workers interested in participating in the study underwent anthropometric measurements and completed questionnaires addressing variables related to demographics, health, lifestyle, work and psychosocial factors. Also on day one, objective measurements of time spent sedentary and in physical activity at work were initiated by equipping the workers with two accelerometers (Actigraph GT3X, ActiGraph LLC, Florida, USA) and a diary for noting working hours. On day four, workers returned the measurement equipment and the diary. Approximately 80% of the workers declared that the objective measurements were collected during typical working days.

Occupational time spent sedentary and in physical activity

Sedentary behavior and physical activity were analyzed using the custom-made Acti4 software according to established procedures (34, 35). The software identifies a number of different activity types, as well as the gross body posture. For the present study, we merged periods of sitting and lying into “sedentary time” and collapsed periods with any type of physical activity into one category, ie, “physical activity”. Thus physical activity is defined to occur whenever the worker is not sitting or lying.

Time spent sedentary and in physical activity was averaged for each specific worker across all working periods with valid measurement data. A working period was considered valid if it comprised ≥ 4 hours of work, and corresponded to $\geq 75\%$ of that individual’s self-reported average working time per day. Workers with at least one valid work day were included in further analyses.

Predictors

The predictors used for modelling in this study were

selected a priori from the questionnaire based on (i) whether they would likely predict time spent sedentary or in physical activity according to previous studies (26, 29, 30, 32, 36–38), (ii) whether they are commonly available in large epidemiological studies and surveys, and (iii) whether they showed a large relative dispersion between workers in our material. Based on these criteria, we arrived at including self-reported information on age, gender, body mass index (BMI), job type, OST, and OPA. These predictors are described in detail in Appendix B (www.sjweh.fi/data_repository.php). Selecting predictors a priori without knowing their relationship to the outcomes is a recommended approach in modern statistical literature as it reduces the risk of capitalizing on chance and arriving at spurious relationships between predictors and outcomes (39–41).

Statistical analyses

All predictors were treated as continuous variables except for gender, job type, and OPA which were treated as categorical variables. Statistical operations were performed using the R software package “rms” (42).

The six predictors (cf. Appendix B) were modeled together against each objectively measured exposure using least-square linear regression analyses to develop a full prediction model. The available degrees of freedom for statistical analysis in this study were sufficient to allow inclusion of all six variables.

The sensitivity of the full model to selection of predictors was analyzed by running models removing one predictor at a time and observing the resulting change in the explained variance.

In addition, single-predictor models were developed based only on those predictors that directly focused on sedentary behavior and/or physical activity. Thus, two models were developed using least-square linear regression analyses, one based on self-reported information on OST (variable #5 in Appendix B) and another based on self-reported OPA (variable #6 in Appendix B). The performance of the resulting full and single-predictor models was evaluated by the R^2 adjusted and the mean squared error (MSE) of estimation. The residuals of the models were examined for normal distribution and homoscedasticity.

The expected ability of the full and single-predictor models in predicting objectively measured exposures in new datasets was estimated using a bootstrap resampling procedure (43). Five-hundred bootstrapped virtual datasets were drawn with replacement from the source population and were of the same size. For each virtual dataset, we fitted a model including the same predictors as in the original model. This re-fitted model was then applied to the source dataset, and the model fit parameters were compared to the parameters obtained in the original fit.

The differences, ie, the “optimism” of the original model, were averaged across all bootstrap repeats and used as an overall measure of optimism, reflecting the extent to which the original model capitalized on chance (44).

Results

Descriptives of the participating workers are shown in table 1. Of the 214 workers, most were engaged in manufacturing (27%) and least in mobile plant operations (5%).

In total, 4357 hours of Actigraph measurements were collected. On average, 38.1% of the working time was spent sedentary and the remaining in physical activity (table 2).

Table 3 shows the resulting coefficients of both the full and the single-predictor models estimating objectively measured sedentary time at work. Since time spent sedentary and in physical activity as defined in the present study are complementary variables (ie, they add up to 100% of the working time) we focus on presenting, in detail, only the results of modelling sedentary time at work. Results pertaining to the prediction model for time spent in physical activity are presented in Appendix C (www.sjweh.fi/data_repository.php).

Table 1. Descriptive statistics of the predictors considered for inclusion in the models estimating objectively measured time spent sedentary and in physical activity at work among blue-collar workers (N=214). [OPA=occupational physical activity; OST=occupational sedentary time; Min=minimum; Max=maximum; SD=standard deviation; BMI=body mass index]

Variables	N	%	Mean	Min	Max	SD
Age (years)	214		44.5	21	65	9.9
Females	87	41				
BMI (kg/m ²)	214		26.4	17.4	42.8	5.0
Job groups						
Health care	16	7				
Assembling	33	15				
Cleaning	32	15				
Construction	39	18				
Manufacturing	58	27				
Garbage collection	25	12				
Mobile plant operations	11	5				
OST [1=almost all the time, 6=never]	194		4.1	1	6	1.1
OPA						
Category 1 ^a	18	9				
Category 2 ^b	23	12				
Category 3 ^c	99	50				
Category 4 ^d	59	30				

^a Mostly sedentary work that does not require strenuous physical activity.

^b Mostly work while standing or walking but not requiring strenuous physical activity.

^c Work while standing or walking with some lifting and carrying.

^d Heavy or fast moving work that is physically strenuous.

Table 2. Descriptives of objectively measured time spent sedentary and in physical activity at work among the studied 214 blue-collar workers. [Min=minimum; Max=maximum; SD=standard deviation]

Variables	Mean	SD ^a	Min	Max	Variance between subjects ^b	Variance between days ^b
Measurement days per worker	2.3	0.9	1	5		
Sedentary time (%)	38.1	18.5	5.9	79.8	320.9	59.3
Physical activity time (%)	61.8	18.7	17.9	94.0	327.4	60.5

^b Calculated using a standard one-way random effects model.

The full model predicted 63% of the variance. Older age and higher BMI were found to be significant predictors of less objectively measured sedentary time at work. Additionally, workers who reported more OPA (answer categories 2–4 of the single four-graded question on OPA, variable #6, cf. Appendix B) were exposed to less objectively measured sedentary time at work compared to those reporting that their work was mostly sedentary and did not require strenuous physical activity (category 1 of the OPA question). Assemblers, garbage collectors and manufacturing workers had markedly less sedentary time at work than mobile plant operators. When predictors were removed from the models one at a time, explained variance was reduced between 1% (gender) and 18% (job group; cf. table 3).

The developed full model (table 3) can be used to predict sedentary time at work for any worker characterized by some specific combination of the predictor variables in the model. For example, a 53-year-old female assembler, with a BMI of 29.1 kg/m², who has reported “mostly sedentary work that does not require strenuous physical activity” (response category 1 of the OPA question), and stated her sedentary time to be “almost all the time”, is predicted by the model to be sedentary 67.8% of her time during working hours according to objective measurement.

Figure 1 shows predicted versus objectively measured values for time spent sedentary and in physical activity at work.

The single-predictor model predicting objectively measured sedentary time at work only from self-reported OPA, showed an explained variance, R² adjusted, of 21% (MSE 260.0; table 3). The other single-predictor model based on the worker’s self-reported OST explained 38% of the variance of objectively measured sedentary time at work (MSE=212.4; table 3).

Internal validation using bootstrapping

The bootstrap validation of the full model revealed an optimism of 5% in explained variance, R². When models developed from the virtual bootstrap datasets were used on the source data, MSE was 136.1, as compared

Table 3. Models predicting objectively measured sedentary time at work on the basis of answers to questions normally available in large epidemiological studies and surveys. [B=regression coefficient; 95% CI=95% confidence interval; R²_{adj}=coefficient of determination (explained variance) adjusted for the number of terms in the model; OPA=occupational physical activity; OST=occupational sedentary time; MSE=mean squared error].

Predictor	B	95% CI	R ² _{adj} ^a (%)	MSE
Single-predictor model based on OST (variable #5 in Appendix B)				38 212.4
Intercept	79.00	71.22–86.77		
OST (1–6)	-10.02	-11.85– -8.20		
Single-predictor model based on OPA (variable #6 in Appendix B)				21 260.0
Intercept	62.28	54.71–69.85		
OPA category (yes=1; no=0)				
2 ^b	-33.57	-43.68– -23.46		
3 ^c	-28.58	-36.81– -20.35		
4 ^d	-22.08	-30.73– -13.43		
Full model				63 117.0
Intercept	110.73	94.84–126.63		
Age (years)	-0.22	-0.40– -0.03		62
Gender (Female=1; male=0)	-4.23	-9.62– -1.16		62
OPA category (yes=1; no=0)				
2 ^b	-15.03	-24.05– -6.01		60
3 ^c	-15.65	-24.16– -7.13		
4 ^d	-11.19	-20.40– -1.98		
BMI (kg/m ²)	-0.38	-0.77– -0.00		61
Occupational sedentary time (1–6)	-6.46	-8.27– -4.66		51
Job group (yes=1; no=0)				
Health services	-7.18	-18.27– -3.91		45
Assembling	-9.68	-19.94– -0.59		
Cleaning	-5.88	-16.15– -4.38		
Construction	-5.42	-14.12– -3.28		
Manufacturing	-24.94	-33.64– -16.24		
Garbage collection	-11.92	-21.40– -2.44		

^a R²_{adj}=explained variance in a model comprising all predictor variables but the one represented by the corresponding row. Note that OPA categories and job groups were removed *en bloc*.

^b Mostly work while standing or walking but not requiring strenuous physical activity.

^c Work while standing or walking with some lifting and carrying.

^d Heavy or fast moving work that is physically strenuous.

to 117.0 in the original fit (cf. table 3). The bootstrap validation of the single-predictor models based on self-reported answers on OPA or OST revealed an optimism of 2% and 0%, respectively, in explained variance when predicting objectively measured sedentary time at work. The corresponding MSE were 268.5 (original fit 260.0) and 215.2 (original fit 212.4).

Discussion

This study aimed at developing and evaluating prediction models for estimating time spent sedentary or in physical activity during working hours by self-reported variables that would normally be available in large epi-

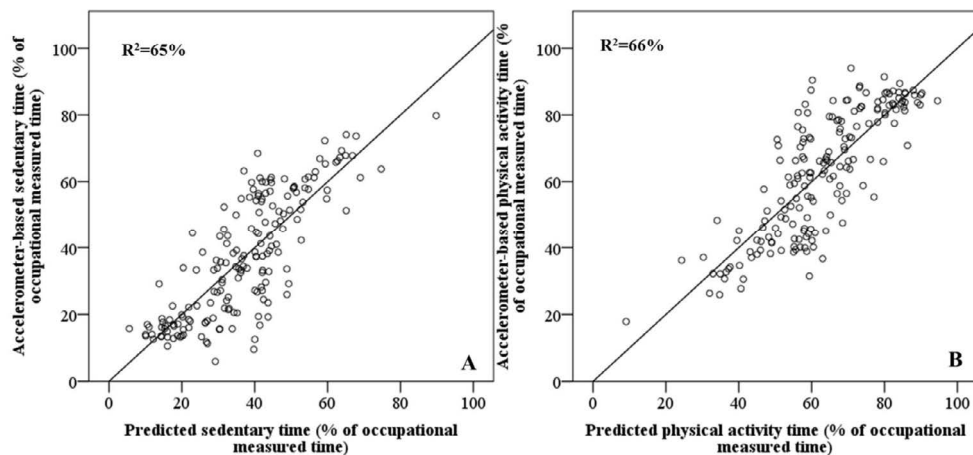


Figure 1. Objectively measured versus predicted occupational time spent sedentary (A; illustrating the full model in table 3) and in physical activity (B; illustrating the complementary full model in Appendix C). Line of identity is included in the diagrams.

demiological studies and surveys. Among blue-collar workers, a full model explained 63% (adjusted R^2) of the variance of objectively measured time spent sedentary, and since we defined physical activity to occur whenever the worker was not sedentary, our model for predicting physical activity had the same performance. Bootstrap validation suggested that this performance was somewhat optimistic; the expected adjusted R^2 when using the model in new datasets would be 57%. Single-predictor models based only on self-reported OPA or OST explained 21% and 38%, respectively, of the variance in objectively measured exposures. The single-predictor models were very stable according to the bootstrap validation.

This study is novel in predicting objectively measured time spent sedentary or in physical activity at work from self-reported information that would normally be available in large epidemiological studies and surveys not specifically designed to predict these exposures. The performance of the full model based on age, job group, BMI, and self-reported OST and OPA was similar to the best performances of previous questionnaires on occupational sedentary and physical activity (45–47). The performance of our model was even better than previously developed models using a customized set of variables to predict self-reported OST or OPA (26, 29, 30). Also, these previous models produce estimates of self-reported OST or OPA, and thus do not adjust for the bias present in these self-reports, relative to objectively measured data. When used in investigations on the effects of sedentary behavior or physical activity on different health outcomes, these models may therefore lead to biased associations and misleading estimates, eg, of the health effects of being sedentary for a particular proportion of the working day. In producing estimates of objectively measured exposure our models avoid this fallacy.

In our full model, age was observed to be negatively associated with sedentary time at work, which

corroborates many previous studies (26, 30, 32). BMI also tended to be negatively associated with time spent sedentary, and this contradicts previous studies (26, 29). However these studies were conducted on the general working population (26, 29) and the models predicted self-reported OST, not objectively measured time as in our study (26, 29). More studies are needed to verify the observed negative association of BMI with objectively measured sedentary time at work, and to disclose whether this association is specific to blue-collar workers.

The predictor “job group” contributed substantially to explain variance in objective exposure, as shown by the considerable decrease in R^2 when removing this particular predictor from the full model (table 3). Job group has previously proven to be an important predictor of sedentary behavior and physical activity (26). However, this previous study categorized participants mainly as white- or blue-collar workers and did not use a detailed occupational classification. Thus, our finding that even job group within the segment of blue-collar workers appears to be an important predictor of sedentary time and physical activity is novel, and encourages further research into the performance of this predictor even for other blue-collar jobs.

As expected, workers reporting more OST were also, on average, more sedentary according to objective measurements, as predicted by our full model. Similarly, workers reporting more OPA (categories 2–4 of the OPA question) generally had less objectively measured sedentary time at work, compared to those reporting that their work was mostly sedentary and did not require strenuous physical activity (category 1 of the OPA question). These results indicate that self-reported OST and OPA do have some potential to predict objective exposures. We did, indeed, find that a fair prediction of objectively measured sedentary time at work could be obtained even using single-predictor models based on a single question about

either OST or OPA (cf. table 3). These results point to a similar or better predictive ability of these self-reported data than found in previous studies addressing sedentary behavior and physical activity (48, 49). Still, answers to these questions cannot explain the major part of the variance of the corresponding objective measures. For the single item measuring OST, one explanation may be that response categories are defined using ambiguous terms such as “almost” and “rarely/very little”, which workers may find difficult to interpret. Answers to the OPA question could only explain 21% of the variance in objectively measured exposures. One reason may be that the four response categories of the OPA question do not fully cover the range of physical activities occurring during work. For example, some workers may have a job comprising considerable sedentary time, but also occasional periods of high physical load; mixed exposures are not clearly captured by any specific response alternative to the OPA question. Notably, this ambiguity may be present even though we slightly modified the answer categories for the OPA question compared to its original form (50) so as to pursue better clarity. For example, response category 1 in the original version (“predominantly sitting”) use the term “sitting” without further specification. This may lead to confusion among workers responding to this question, and the modified OPA answer attempts to be more specific in saying, “Mostly sedentary work that does not require strenuous physical activity”. In response category 4, the term “heavy manual work” used for the original OPA question was replaced with “heavy or fast moving work that is physically strenuous”, which we believe is more explanatory. Due to these slight discrepancies, our prediction models may perform differently if they are applied to answers on the original OPA question.

For a prediction model to be useful, it should not only perform well for the dataset on which it was developed, but also for new datasets. We could not perform a genuine external validation of our prediction models since a new similar dataset was not available. In lack of new data for testing the models’ performance, we used an alternative validation technique, i.e., internal bootstrap validation, which has been recommended in statistical literature on modelling (40, 44, 51). The bootstrap validation suggested a 5% optimism in variance explained by the full models predicting time spent sedentary and in physical activity at work. Thus, a considerable proportion of exposure variance could be explained by the full models even after adjusting for optimism. We therefore consider our models to be useful in other datasets, even when taking into consideration that bootstrap validation may result in a too-positive impression of model performance since it inherently reflects the structure of the source data. Also, as our objective measurements were collected for a limited number of days, which leads to uncertainty in the resulting “true” mean exposure estimates, we would expect

our models to perform even better in predicting mean exposures for longer periods of time, for which sampling variance will be less pronounced.

Our definition and operationalization of sedentary time is consistent with the majority of previous research on sedentary behavior (52, 53). However, defining physical activity as being strictly complementary to sedentary time (ie, equivalent to “non-sedentary time”) deviates from some previous studies (54, 55). This reflects an inconsistency in the literature regarding how to define physical activity. For example, Smith, Hamer (55) defined physical activity on basis of a step count while Buman, Hekler (54) used thresholds based on accelerometer counts. Due to this discrepancy in the operational definition of physical activity, we emphasize that our models are valid only when adopting equivalent definitions of physical activity.

Practical implication of the findings and future research

Our full prediction models will be particularly useful in future studies in which data can be collected on all predictors. The models may also, with some hesitation, be used to obtain post-hoc estimates of objectively measured time spent sedentary and in physical activity at work in existing studies if data on the predictor variables included in our models are available. Danish national surveys of working life regularly collect self-reported information on all predictor variables used in our model. In the Danish national survey on work and health conducted in 2010 (56), data were collected on age, BMI, gender, OST, OPA, and job group using similar questions as in the present study. Thus, we believe that our full models have the potential to be used in a retrospect reconsideration of data on exposures and occupational risks, at least in Danish national surveys. We emphasize that the prediction models developed in the present study are specific to the source population of blue-collar workers, and we recommend future studies to test our models in other occupations and, if necessary, develop adjusted models that fit those occupations better. We also encourage studies testing our models for use in the general population.

Additional predictors that we did not include in our models may have the ability to predict sedentary time and physical activity, for instance psychosocial variables (26). Thus, we encourage investigations into the contents of other large-scale survey data materials to disclose whether they do, indeed, include the predictors used in our models (or subsets thereof) and/or additional potential predictors, for instance psychosocial variables. To this end, variables that may strongly depend on the character of the job, psychosocial factors being one example, should be included with caution, due to their potentially strong correlation with occupation; and a possible predictive ability should be interpreted with this correlation in mind.

We encourage efforts to develop new, short, yet reliable questionnaires assessing occupational sedentary behavior and physical activity, and verify if they can predict “true” exposures better than the self-reported OPA or OST used in our study. If so, they could be attractive for future large scale population surveys. In this line of development, we encourage paying consideration to the trade-off between resources required to collect predictors and the eventual performance of the model. We did not in the present study compare costs associated with collecting exposure data directly by objective measurements versus costs of developing prediction models using questionnaire data, but we emphasize that a major rationale for modeling exposures in future studies would be that they deliver more exposure data at a lower cost than objective measurements, and thus represent a favorable trade-off between cost and performance (57, 58). We also emphasize the need for validating the performance of any future model in new datasets, either using a genuinely new sample or by internal bootstrap validation techniques; this has very rarely been attempted in previous modelling studies (41).

Concluding remarks

This study showed that full prediction models based on self-reported information that would normally be available in large epidemiological studies and surveys could predict objectively measured time spent sedentary and in physical activity at work with a reasonably good accuracy. Internal validation of the prediction models indicated that performance would decrease slightly if they were used in new datasets from the studied occupations, but that they may still be useful for other populations of blue-collar workers, with due caution. Single-predictor models using information specifically addressing sedentary behavior and physical activity showed lower performance than full models, but still offered an attractive opportunity to predict objectively measured exposures. We suggest that our models may be used to revisit previous studies based only on self-reports, as well as an inspiration for designing future large epidemiological studies and surveys addressing sedentary behavior and physical activity on the basis of self-reported information.

Conflict of interest

The authors declare no conflict of interest.

Financial support

The Danish Work Environment Research Fund supported this work (grant number 20130069161/9).

References

1. Krause N, Brand RJ, Kaplan GA, Kauhanen J, Malla S, Tuomainen TP, et al. Occupational physical activity, energy expenditure and 11-year progression of carotid atherosclerosis. *Scand J Work Environ Health*. 2007;33(6):405–24. <http://dx.doi.org/10.5271/sjweh.1171>.
2. Holtermann A, Hansen JV, Burr H, Sjøgaard K, Sjøgaard G. The health paradox of occupational and leisure-time physical activity. *Br J Sports Med*. 2012;46(4):291–5. <http://dx.doi.org/10.1136/bjsm.2010.079582>.
3. Det Nationale Forskningscenter for Arbejdsmiljø. Arbejdsmiljø og helbred i Danmark 2012 [Work Environment and Health Survey in Denmark 2012] Denmark2012 [cited 2014 21 July]. Available from: <http://www.arbejdsmiljoforskning.dk/~media/Projekter/AH2012/AH-2012-spoergeskema-endelig-version.pdf>.
4. Lund T, Labriola M, Christensen KB, Bultmann U, Villadsen E. Physical work environment risk factors for long term sickness absence: prospective findings among a cohort of 5357 employees in Denmark. *Br Med J*. 2006;332(7539):449–51. <http://dx.doi.org/10.1136/bmj.38731.622975.3A>.
5. Sell L, Bultmann U, Rugulies R, Villadsen E, Faber A, Sogaard K. Predicting long-term sickness absence and early retirement pension from self-reported work ability. *Int Arch Occup Environ Health*. 2009;82(9):1133–8. <http://dx.doi.org/10.1007/s00420-009-0417-6>.
6. Holtermann A, Burr H, Hansen JV, Krause N, Sogaard K, Mortensen OS. Occupational physical activity and mortality among Danish workers. *Int Arch Occup Environ Health*. 2012;85(3):305–10. <http://dx.doi.org/10.1007/s00420-011-0668-x>.
7. Harari G, Green MS, Zelber-Sagi S. Combined association of occupational and leisure-time physical activity with all-cause and coronary heart disease mortality among a cohort of men followed-up for 22 years. *Occup Environ Med*. 2015;72(9):617–24. <http://dx.doi.org/10.1136/oemed-2014-102613>.
8. Chau JY, van der Ploeg HP, Merom D, Chey T, Bauman AE. Cross-sectional associations between occupational and leisure-time sitting, physical activity and obesity in working adults. *Prev Med*. 2012;54. <http://dx.doi.org/10.1016/j.ypmed.2011.12.020>.
9. Clemes SA, Patel R, Mahon C, Griffiths PL. Sitting time and step counts in office workers. *Occup Med (Lond)*. 2014;64. <http://dx.doi.org/10.1093/occmed/kqt164>.
10. Varela-Mato V, Yates T, Stensel DJ, Biddle SJ, Clemes SA. Time spent sitting during and outside working hours in bus drivers: A pilot study. *Prev Med Rep*. 2016;3:36–9. <http://dx.doi.org/10.1016/j.pmedr.2015.11.011>.
11. van Uffelen JGZ, Wong J, Chau JY, van der Ploeg HP, Riphagen I, Gilson ND, et al. Occupational sitting and health risks: A systematic review. *Am J Prev Med*. 2010;39(4):379–88. <http://dx.doi.org/10.1016/j.amepre.2010.05.024>.

12. Lis AM, Black KM, Korn H, Nordin M. Association between sitting and occupational LBP. *Eur Spine J*. 2007;16(2):283–98. <http://dx.doi.org/10.1007/s00586-006-0143-7>.
13. Gupta N, Hallman DM, Mathiassen SE, Aadahl M, Jorgensen MB, Holtermann A. Are temporal patterns of sitting associated with obesity among blue-collar workers? A cross sectional study using accelerometers. *BMC Public Health*. 2016;16(1):148. <http://dx.doi.org/10.1186/s12889-016-2803-9>.
14. Terwee CB, Mokkink LB, van Poppel MN, Chinapaw MJ, van Mechelen W, de Vet HC. Qualitative attributes and measurement properties of physical activity questionnaires: a checklist. *Sports Med*. 2010;40(7):525–37. <http://dx.doi.org/10.2165/11531370-000000000-00000>.
15. Clays E, De Bacquer D, Van Herck K, De Backer G, Kittel F, Holtermann A. Occupational and leisure time physical activity in contrasting relation to ambulatory blood pressure. *BMC Public Health*. 2012;12:1002. <http://dx.doi.org/10.1186/1471-2458-12-1002>.
16. Arbejdsmiljø og helbred i Danmark. Dataset, Fysisk krav: Arbejdsmiljø og helbred i Danmark 2012 [Dataset, physical demands: Work Environment and Health Survey in Denmark 2012]. National Research Centre for the Working Environment; 2012.
17. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *Int J Behav Nutr Phys Act*. 2008;5:56. <http://dx.doi.org/10.1186/1479-5868-5-56>.
18. Plasqui G, Westerterp KR. Physical activity assessment with accelerometers: an evaluation against doubly labeled water. *Obesity (Silver Spring)*. 2007;15(10):2371–9. <http://dx.doi.org/10.1038/oby.2007.281>.
19. Kelly LA, McMillan DG, Anderson A, Fippinger M, Fillerup G, Rider J. Validity of actigraphs uniaxial and triaxial accelerometers for assessment of physical activity in adults in laboratory conditions. *BMC Med Phys*. 2013;13(1):5. <http://dx.doi.org/10.1186/1756-6649-13-5>.
20. Wolin KY, Heil DP, Askew S, Matthews CE, Bennett GG. Validation of the International Physical Activity Questionnaire-Short among Blacks. *J Phys Act Health*. 2008;5(5):746–60.
21. Bringolf-Isler B, Mader U, Ruch N, Kriemler S, Grize L, Braun-Fahrlander C. Measuring and validating physical activity and sedentary behavior comparing a parental questionnaire to accelerometer data and diaries. *Pediatr Exercise Science*. 2012;24(2):229–45.
22. Trask C, Mathiassen SE, Jackson J, Wahlstrom J. Data processing costs for three posture assessment methods. *BMC Med Res Methodol*. 2013;13:124. <http://dx.doi.org/10.1186/1471-2288-13-124>.
23. Tucker JM, Welk G, Nusser SM, Beyler NK, Dziewaltowski D. Estimating minutes of physical activity from the previous day physical activity recall: validation of a prediction equation. *J Phys Act Health*. 2011;8(1):71–8.
24. Saint-Maurice PF, Welk GJ, Beyler NK, Bartee RT, Heelan KA. Calibration of self-report tools for physical activity research: the Physical Activity Questionnaire (PAQ). *BMC Public Health*. 2014;14:461. <http://dx.doi.org/10.1186/1471-2458-14-461>.
25. Beyler N, Fuller W, Nusser S, Welk G. Predicting objective physical activity from self-report surveys: a model validation study using estimated generalized least-squares regression. *J Appl Stat*. 2014;42(3):555–65. <http://dx.doi.org/10.1080/0264763.2014.978271>.
26. De Cocker K, Duncan MJ, Short C, van Uffelen JG, Vandelanotte C. Understanding occupational sitting: prevalence, correlates and moderating effects in Australian employees. *Prev Med*. 2014;67. <http://dx.doi.org/10.1016/j.ypmed.2014.07.031>.
27. Wallmann-Sperlich B, Bucksch J, Schneider S, Froboese I. Socio-demographic, behavioural and cognitive correlates of work-related sitting time in German men and women. *BMC Public Health*. 2014;14:1259. <http://dx.doi.org/10.1186/1471-2458-14-1259>.
28. Wallmann-Sperlich B, Froboese I. Physical activity during work, transport and leisure in Germany--prevalence and socio-demographic correlates. *PLoS ONE*. 2014;9(11):e112333. <http://dx.doi.org/10.1371/journal.pone.0112333>.
29. Hadgraft NT, Lynch BM, Clark BK, Healy GN, Owen N, Dunstan DW. Excessive sitting at work and at home: Correlates of occupational sitting and TV viewing time in working adults. *BMC Public Health*. 2015;15:899. <http://dx.doi.org/10.1186/s12889-015-2243-y>.
30. Bennie JA, Pedisic Z, Timperio A, Crawford D, Dunstan D, Bauman A, et al. Total and domain-specific sitting time among employees in desk-based work settings in Australia. *Aust N Z J Public Health*. 2015;39(3):237–42. <http://dx.doi.org/10.1111/1753-6405.12293>.
31. Kwak L, Proper KI, Hagströmer M, Sjöström M. The repeatability and validity of questionnaires assessing occupational physical activity—a systematic review. *Scand J Work Environ Health*. 2011;37(1):6–29. <http://dx.doi.org/10.5271/sjweh.3085>.
32. Wallmann-Sperlich B, Bucksch J, Schneider S, Froboese I. Socio-demographic, behavioural and cognitive correlates of work-related sitting time in German men and women. *BMC Public Health*. 2014;14(1):1–10. <http://dx.doi.org/10.1186/1471-2458-14-1259>.
33. Gupta N, Jensen BS, Sogaard K, Carneiro IG, Christiansen CS, Hanisch C, et al. Face validity of the single work ability item: comparison with objectively measured heart rate reserve over several days. *Int J Env Res Public Health*. 2014;11(5):5333–48. <http://dx.doi.org/10.3390/ijerph110505333>.
34. Gupta N, Stordal CC, Hallman D, Korshøj M, Gomes CI, Holtermann A. Is objectively measured sitting time associated with low back pain? A cross-sectional investigation in the NOMAD study. *PLoS ONE*. 2015;10(3):e0121159. <http://dx.doi.org/10.1371/journal.pone.0121159>.
35. Skotte J, Korshoj M, Kristiansen J, Hanisch C, Holtermann A. Detection of physical activity types using triaxial

- accelerometers. *J Phys Act Health*. 2014;11(1):76–84. <http://dx.doi.org/10.1123/jpah.2011-0347>.
36. Duncan MJ, Short C, Rashid M, Cutumisu N, Vandelanotte C, Plotnikoff RC. Identifying correlates of breaks in occupational sitting: a cross-sectional study. *Build Res Inf*. 2015;43(5):646–58. <http://dx.doi.org/10.1080/09613218.2015.1045712>.
 37. Vandelanotte C, Duncan MJ, Short C, Rockloff M, Ronan K, Happell B, et al. Associations between occupational indicators and total, work-based and leisure-time sitting: a cross-sectional study. *BMC Public Health*. 2013;13:1110. <http://dx.doi.org/10.1186/1471-2458-13-1110>.
 38. Wallmann-Sperlich B, Froboese I. Physical activity during work, transport and leisure in Germany - prevalence and socio-demographic correlates. *PLoS ONE*. 2014;9. <http://dx.doi.org/10.1371/journal.pone.0112333>.
 39. Lavoue J, Droz PO. Multimodel inference and multimodel averaging in empirical modeling of occupational exposure levels. *Ann Occup Hyg*. 2009;53(2):173–80. <http://dx.doi.org/10.1093/annhyg/men085>.
 40. Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. NY,USA: Springer; 2001. <http://dx.doi.org/10.1007/978-1-4757-3462-1>.
 41. Heiden M, Mathiassen SE, Garza J, Liv P, Wahlstrom J. A Comparison of Two Strategies for Building an Exposure Prediction Model. *Ann Occup Hyg*. 2016;60(1):74–89.
 42. Harrell F. *Regression Modeling Strategies*. R package version 4.4-2; 2016.
 43. Suarathana E, Vergouwe Y, Moons KG, de Monchy J, Grobbee D, Heederik D, et al. A diagnostic model for the detection of sensitization to wheat allergens was developed and validated in bakery workers. *J Clin Epidemiol*. 2010;63(9):1011–9. <http://dx.doi.org/10.1016/j.jclinepi.2009.10.008>.
 44. Efron B. Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation. *J Amer Statist Assoc*. 1983;78(382):316–31. <http://dx.doi.org/10.1080/01621459.1983.10477973>.
 45. Helmerhorst HJ, Brage S, Warren J, Besson H, Ekelund U. A systematic review of reliability and objective criterion-related validity of physical activity questionnaires. *Int J Behav Nutr Phys Act*. 2012;9:103. <http://dx.doi.org/10.1186/1479-5868-9-103>.
 46. Chau JY, Van Der Ploeg HP, Dunn S, Kurko J, Bauman AE. Validity of the occupational sitting and physical activity questionnaire. *Med Sci Sports Exerc*. 2012;44(1):118–25. <http://dx.doi.org/10.1249/MSS.0b013e3182251060>.
 47. Kwak L, Proper KI, Hagströmer M, Sjöström M. The repeatability and validity of questionnaires assessing occupational physical activity--a systematic review. *Scand J Work Environ Health*. 2011;37(1):6–29. <http://dx.doi.org/10.5271/sjweh.3085>.
 48. Emaus A, Degerstrom J, Wilsgaard T, Hansen BH, Dieli-Conwright CM, Furberg AS, et al. Does a variation in self-reported physical activity reflect variation in objectively measured physical activity, resting heart rate, and physical fitness? Results from the Tromso study. *Scand J Public Health*. 2010;38(5 Suppl):105–18. <http://dx.doi.org/10.1177/1403494810378919>.
 49. Wiktorin C, Karlqvist L, Winkel J. Validity of self-reported exposures to work postures and manual materials handling. Stockholm MUSIC I Study Group. *Scand J Work Environ Health*. 1993;19(3):208–14. <http://dx.doi.org/10.5271/sjweh.1481>.
 50. Saltin B, Grimby G. Physiological analysis of middle-aged and old former athletes. Comparison with still active athletes of the same ages. *Circulation*. 1968;38(6):1104–15. <http://dx.doi.org/10.1161/01.CIR.38.6.1104>.
 51. Steyerberg EW, Harrell FE, Jr. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol*. 2016;69:245–7.
 52. Wijndaele K, I DEB, Godino JG, Lynch BM, Griffin SJ, Westgate K, et al. Reliability and validity of a domain-specific last 7-d sedentary time questionnaire. *Med Sci Sports Exerc*. 2014;46(6):1248–60. <http://dx.doi.org/10.1249/MSS.0000000000000214>.
 53. Sedentary Behaviour Research N. Letter to the editor: standardized use of the terms “sedentary” and “sedentary behaviours”. *Appl Physiol Nutr Metab*. 2012;37(3):540–2. <http://dx.doi.org/10.1139/h2012-024>.
 54. Buman MP, Hekler EB, Haskell WL, Pruitt L, Conway TL, Cain KL, et al. Objective light-intensity physical activity associations with rated health in older adults. *Am J Epidemiol*. 2010;172(10):1155–65. <http://dx.doi.org/10.1093/aje/kwq249>.
 55. Smith L, Hamer M, Ucci M, Marmot A, Gardner B, Sawyer A, et al. Weekday and weekend patterns of objectively measured sitting, standing, and stepping in a sample of office-based workers: the active buildings study. *BMC Public Health*. 2015;15:9. <http://dx.doi.org/10.1186/s12889-014-1338-1>.
 56. Det Nationale Forskningscenter for Arbejdsmiljø. *Arbejdsmiljø og helbred i Danmark 2010 [Work Environment and Health Survey in Denmark 2010] Denmark2010 [cited 2014 21 July]*. Available from: <http://www.arbejdsmiljoforskning.dk/~media/Projekter/AH2012/AH-2012-spoergeskema-enderlig-version.pdf>.
 57. Trask C, Mathiassen SE, Wahlstrom J, Forsman M. Cost-efficient assessment of biomechanical exposure in occupational groups, exemplified by posture observation and inclinometry. *Scand J Work Environ Health*. 2014;40(3):252–65. <http://dx.doi.org/10.5271/sjweh.3416>.
 58. Trask C, Teschke K, Village J, Chow Y, Johnson P, Luong N, et al. Measuring low back injury risk factors in challenging work environments: an evaluation of cost and feasibility. *Am J Ind Med*. 2007;50(9):687–96. <http://dx.doi.org/10.1002/ajim.20497>.
 59. International Labour Organisation. *ISCO: International Standard Classification of Occupations [cited 2013 June 20]*. Available from: <http://www.ilo.org/public/english/bureau/stat/isco/isco08/index.htm>.

Received for publication: 19 August 2015