This is a non-final version of an article published in final form in *Medicine and Science in Sports and Exercise*, 45 (5), 2013, 1004-1011.

Is Measurement Error Altered by Participation in a Physical Activity Intervention?

Elisabeth Winkler[1], Lauren Waters[1], Elizabeth Eakin[1], Brianna Fjeldsoe[1], Neville Owen[2§], Marina Reeves[1]

[1] The University of Queensland, School of Population Health, Cancer Prevention Research Centre, Brisbane, Australia.

[2] Baker IDI Heart and Diabetes Institute, Behavioural Epidemiology Laboratory, Melbourne, Australia.

[§]Corresponding author:

Baker IDI Heart and Diabetes Institute, Behavioural Epidemiology Laboratory, 75 Commercial road, Melbourne Victoria 3004, Australia

Email: Neville.Owen@bakeridi.edu.au; Phone: +61 3 8532 1874; Fax: +61 3 8532 1100

Running Title: Is measurement error altered by intervention?

Conflicts of Interest and Source of Funding: This study was supported by grants from the Australian National Health and Medical Research Council (NHMRC) and the Diabetes Australia Research Trust. Authors were supported by Queensland Health core infrastructure funding (Winkler); a National Heart Foundation scholarship (Waters); and, NHMRC

Fellowships (Eakin, Owen and Reeves). No other sources of funding or conflicts of interest were declared by the authors. The trial is registered with the Australian Clinical Trials Registry #ACTRN12608000203358.

ABSTRACT

Purpose: There is no 'gold standard' measure for moderate to vigorous physical activity (MVPA) (11); some error is inherent to self-report and device-based measures. Few studies have examined agreement between self-report and device-based measures in the intervention trial context or whether the difference between measures is influenced by intervention participation.

Methods: MVPA was measured at baseline and 6-months by Active Australia survey (AAS) and GT1M accelerometer (≥ 1952 counts/minute) in the intervention (*n* =135) and usual care control (*n* =141) participants of a randomized trial targeting weight loss by MVPA increases and energy intake reductions in adults with type 2 diabetes. Agreement, for each group at each assessment, was examined using the Bland-Altman approach and regression based modelling. As the differences between MVPA measures varied with average values ([AAS + GT1M]/ 2), they were examined as a percentage of average physical activity. *T*-tests were used to assess unadjusted group differences and changes over time. Analysis of covariance models tested intervention effects on measurement error at follow-up, adjusted for baseline.

Results: Agreement worsened, and variability in the difference measures became greater, as the average amount of MVPA increased. Measurement error differed significantly between groups at follow-up (*P* = .010) but not baseline (*P* = .157) and changed significantly within the intervention group (*P* = .001) but not the control group (*P* = .164). There was a statistically significant effect of the intervention on measurement error (*P* = .026).

Conclusions: Measurement error of self-report relative to accelerometer= appeared to be affected by intervention. As measurement error cannot be definitively attributed to self-report or accelerometer, it would be prudent to measure both in future studies.

Key Words: validity, self report, accelerometer, randomized controlled trial

INTRODUCTION

Paragraph Number 1 Physical inactivity is a modifiable risk factor for chronic diseases including type 2 diabetes, cardiovascular disease, and some cancers. The development of effective, broad reaching physical activity interventions is a population health priority and in order to properly evaluate such interventions, valid and reliable measures are needed. There is no 'gold standard' measure for moderate-to-vigorous physical activity (MVPA) (11); interventions most commonly use self-report but increasingly incorporate objective monitoring. A systematic review (17) has shown self-report typically has only low to moderate correlations with measures from monitoring devices, including accelerometers, and can yield much lower or higher estimates of physical activity than monitoring devices. The substantial disagreement between self-report and objective monitoring has implications for estimating intervention effects, as there is reason to suspect measurement error may be differential between intervention and control groups after intervention (22).

Paragraph Number 2 A study of adolescent girls (22) observed measurement error (self-report relative to accelerometer) was differential at follow-up but not at baseline and explored social desirabiltiy bias as a possible reason. Social desirability bias may cause interventions to appear more effective by self-report than they really were if, after intervention, participants who have received an intervention that involves intensive contact with intervention staff are more inclined to report their behavior in a socially desirable manner than controls who have had minimal or no contact with staff. Self-report may alternatively underestimate intervention effects, for example, if certain intervention components (e.g., self-monitoring) increase intervention participants' awareness of their physical activity and thereby selectively improve their reporting accuracy (i.e, reduce their over-reporting) (23, 26). Further, disagreement between self-report and accelerometer may be altered by interventions that increase physical

activity as it is often reported to be worse at higher levels of physical activity than at lower levels (2, 9, 21).

Paragraph Number 3 Whether measurement error is impacted by intervention in adults, and the implications this has for trial outcomes remains to be established. In the context of a behaviorally based weight loss intervention trial (10), we examined the agreement between self-reported and accelerometer-measured MVPA at baseline and 6-months within the control and intervention groups. We further examined variation over time, differences between groups and intervention effects on the difference between the measures of MVPA. We also examined the potential implications of any alteration in measurement error over the course of the intervention by comparing the trial's intervention effects according to self-report and accelerometer.

METHODS

Paragraph Number 4 This study uses data from Living Well with Diabetes (LWWD), a randomized controlled trial of a telephone-delivered weight loss intervention targeting increased physical activity and reductions in energy intake. The study protocol has been reported previously (10). Recruitment of primary care practices commenced in October 2008 with participant recruitment occurring between February 2009 and April 2011. Data collection occurred at baseline and 6 months and is ongoing for 12-, 18- and 24-months. Informed consent was obtained from all participants and the University of Queensland Behavioral and Social Sciences Ethics Review Committee granted ethics approval for the trial.

Paragraph Number 5 Nine general practices were recruited in Logan, a diverse socioeconomic area approximately 30km south of Brisbane, Australia. Within practices, 1407 eligible patients (i.e., diagnosed type 2 diabetes, aged 20 - 75 years, and having a listed

telephone number) were identified using electronic medical records. Patients not excluded by GP screening for contraindications to unsupervised physical activity (*n* = 908) were posted study materials by the GP and were followed up by study staff to determine eligibility and solicit informed consent unless they had declined further contact (*n* = 206). To be eligible, participants needed to be overweight or obese (Body Mass Index, BMI ≥ 25.0 kg/m^2) (28) or inactive (< 5 days per week of ≥ 30 minutes of MVPA) (3) and could not report: current cancer treatment other than endocrine therapy; prior or planned bariatric surgery; use of weight loss medications (e.g., Orlistat) or communication difficulties relevant to receiving the intervention. Of those reached by telephone and deemed eligible (*n* = 420), 302 (71.9%) agreed to participate, completed the baseline assessment and were randomized to receive either Telephone Counseling (*n* = 151) or Usual Care (*n* = 151). Randomization was by the minimization method (1, 19) using the MINIM program (www.sghms.ac.uk/depts/phs/guide/randser.htm), aiming to balance treatment groups across the following prognostic factors (without weighting for importance): gender; age (≥ 55 years); BMI (≥ 40 kg/m^2); HbA1c (≥ 8%); self-reported MVPA (≥ 150 minutes and ≥ 5 days per week); and, diabetes management (insulin or combination therapy / traditional oral hypoglycaemic medications / new diabetes agents, i.e. Exenatide or Sitagliptin / lifestyle alone).

Telephone Counseling Weight Loss Intervention

Paragraph Number 6 Intervention participants were assigned at least 14 calls over the intensive first 6-months of an 18-month telephone-delivered weight loss counseling intervention, described in detail elsewhere (10). Participants were provided with a detailed workbook, pedometer, a self-monitoring ‘tracker’ to record daily physical activity and food intake and detailed feedback after each assessment that highlighted discrepancies between their reported behavior (including physical activity) and the study targets. The physical

activity target was at least 210 minutes of planned, moderate-intensity physical activity per week (≥ 30 minutes per day every day). A further two to three resistance training sessions per week were also recommended.

Usual Care

Paragraph Number 7 Following each assessment, participants in the control group received publically available brochures addressing health behaviors important for diabetes self-management (e.g., losing weight, alcohol consumption) and a letter thanking them for their participation. The letter included brief feedback summarizing their assessment results, including their physical activity levels, but avoided any comparison with recommendations. The attention provided to the control group was designed to minimize attrition.

Data Collection and Processing

Paragraph Number 8 Self-reported physical activity. Self-reported physical activity was assessed by Computer Assisted Telephone Interview (CATI) using the Active Australia Survey (AAS), an eight-item questionnaire that is used extensively in Australian research (4). Participants were asked to report the total time engaged in gardening, walking (for ≥ 10 minutes at a time), moderate- and vigorous-intensity physical activities in the past week. Data were processed according to the AAS protocol (4), but without doubling of the duration of vigorous activity, to facilitate comparability with the accelerometer data. Total weekly MVPA was calculated as the summed durations of walking, moderate-, and vigorous-intensity activities (excluding gardening), first truncating each activity at 840 minutes/week and truncating total MVPA at 1680 minutes/week. In assessing duration of MVPA, the AAS has test-retest reliability that is comparable with the U.S. Behavioral Risk Factor Surveillance Survey and the International Physical Activity Questionnaire (9), has acceptable agreement with these measures (7) and moderate correlation with accelerometer, ranging from 0.29

(95% CI: 0.16, 0.41) (24) to 0.52 (95% CI not reported) (8). Intra-class correlations reported for test-retest range from 0.32 (95% CI: 0.09, 0.52) (16) to 0.64 (95% CI: 0.57, 0.70) (9).

Paragraph Number 9 Accelerometer measurement of physical activity. Actigraph GT1M accelerometers (Actigraph, LLC, Fort Walton Beach, Florida), worn on elasticized belts around participants' waists, positioned on the right mid-axillary line, were used to monitor MVPA objectively. Participants were instructed to wear the accelerometers during all waking hours for seven consecutive days, except during water-based activities (e.g., showering, swimming). The accelerometers were set to collect data in 60-second epochs. Participants reported the times the accelerometer was donned and removed each day in a wearing log. Research staff identified non-wear periods by comparing participants' log data with the precise times movement began and ceased, to overcome limitations in relying exclusively on wear-time algorithms (27) or self-report. Days were considered valid that had at least 10 hours of wear and no implausibly high counts (≥ 20,000 counts per minute; cpm). Accelerometer compliance was good, with almost all participants providing at least four valid days of wear (98.1% at baseline and 97.1% at 6-months). Mean (± Standard Deviation, SD) wear time on valid days was 13.5 ± 1.7 hours/day at baseline and very similar at 6-months (13.7 ± 1.7 hours/day). A program adapted from the National Cancer Institute was used to process and summarise the data (15). For each valid day, MVPA was identified as the number of minutes at or above the commonly used cutpoint of 1952 cpm (12). Average MVPA on valid days was multiplied by 7 to yield a weekly estimate. Sensitivity analyses examined alternative MVPA classifications: one of the lowest (≥ 574 cpm) and highest 3-MET cutpoints for MVPA (≥ 2743 cpm) (25).

Paragraph Number 10 Anthropometric and demographic measures. Anthropometric and demographic data were collected at baseline. BMI was determined from nurse-assessed

height and weight; demographic information was collected via CATI by research staff blinded to study group allocation (10).

Statistical Analyses

Paragraph Number 11 Data analyses were performed in SPSS (IBM SPSS v.20; SPSS) and significance was set at $P < 0.05$. Analyses included the $n = 272$ participants with physical activity data from both self-report and accelerometer at both baseline and 6-months. Agreement of self-report with accelerometer MVPA was assessed using the Bland-Altman approach. Scatterplots depicting the differences between measures (AAS - GT1M) across the average value of the two measures ([AAS + GT1M]/2) (5) were created separately for each group at each time point. As regression (5, 6) showed mean differences and variability both increased significantly ($P < 0.05$) with average values, and log-transformation failed to resolve this issue, the scatterplots show mean differences and 95% Limits of Agreement (LoA) in terms of the regression equations. For this same reason, when examining changes, group differences and intervention effects on measurement error, the difference between the MVPA measures was examined as a percentage of average values (100x[AAS - GT1M]/ [AAS + GT1M]/2). This outcome followed an approximately normal distribution. Unadjusted differences between groups in means (systematic error) were examined using independent samples *T*-tests; Levene's test was used to test differences between groups in variance (random error). Changes over time within groups were examined by paired t-tests.

Paragraph Number 12 Analysis of covariance (ANCOVA) models were used to examine intervention effects in terms of group differences in measurement error at follow-up, adjusted for baseline measurement error. To examine the potential impact of measurement error on results from the intervention, the trial's intervention effects for MVPA by self-report and by accelerometer are reported. While AAS and GT1M data were skewed, change scores (6-months minus baseline) approximated normality. ANCOVA models examined change in

MVPA, adjusted for baseline MVPA. From these models, group differences and the adjusted mean change within groups are reported, with 95% Confidence Intervals (CI). All models showed minimal non-normality and heteroscedascicity. Sensitivity analyses were conducted to examine whether the main study conclusions regarding changes, group differences or intervention effects on measurement error were robust across a wide range of potential cutpoints for MVPA.

RESULTS

Participants

Paragraph Number 13 Characteristics of study participants are presented in Table 1. Trial participants had a mean (± Standard Deviation, SD) age of 58.3 (± 8.6) years, were nearly all either overweight (26.2%) or obese (68.2%), mostly Caucasian (87.4%) and were more commonly male (56.3%) than female. Median duration of diabetes was five years (25^{th}, 75^{th} percentile: 2, 10 years). The trial had a high response rate (72% of those reached and eligible) and participants did not differ from non-participants on most variables (see Table, SDC 1, comparison of study participants with non-participants on demographic, health, and behavioral characteristics), however, participants had significantly higher BMI, shorter diabetes duration, more education, and were under-represented by ex-smokers and over-represented by never smokers. Loss to follow-up was low (12.6% intervention and 7.3% control) and was non-differential. Most characteristics did not differ between those with complete (n = 272) and incomplete data (n = 30) (see Table, SDC 2, comparison of completers with those missing 6-month study outcomes); the only statistically significant differences were for use of insulin (P = 0.023) and for smoking status (P = 0.036), with those on insulin and current smokers tending to lack follow-up data.

INSERT TABLE 1 ABOUT HERE

Agreement of the AAS with GT1M at Baseline and Six Months

Paragraph Number 14 The Bland-Altman plots (Figure 1 a-d) show the agreement of self-reported MVPA (AAS) with accelerometer-measured MVPA (GT1M) at baseline in a) intervention and b) control groups and at 6-months in c) intervention d) control. For each group at each time point, as the average value of the two measures ([AAS + GT1M]/2) increased, the mean difference (AAS - GT1M) and the variability in the differences also increased significantly (all comparisons $P < 0.05$). This can be seen in all four plots in the slope of the lines for the mean differences and 95% LoA, which respectively increase, and widen, as average values increase. The mean differences and 95% LoA indicated substantial differences between self-report and accelerometer, however the plot for the intervention group at follow-up showed smaller mean differences and narrower limits of agreement than the other plots. For example, Panel a (Figure 1) shows that for the intervention group at baseline, we would expect a mean difference of 104.1 minutes/week (i.e. 68.1 + 0.6x60) with 95% LoA of -190.5 to 120.2 minutes/week for those with 60 minutes/week as their average MVPA ([AAS + GTIM]/2) and a much larger mean difference (194.1 minutes/week) and wider LoA (-445.5 to 540.2 minutes/week) in those with much higher physical activity levels (average values of 210 minutes/week). By contrast at follow-up in the intervention group, the mean differences and LoA were much smaller for a given amount of average MVPA, e.g., 19.9 (95% LoA: -187.7, 195.8) minutes/week for those with average values of 60 minutes/week and 49.9 (95% LoA: -292.7, 360.8) minutes/week for those with average values of 210 minutes/week.

INSERT FIGURE 1 ABOUT HERE

Effect of the Intervention on Measurement Error (Difference Between AAS and GT1M)

Paragraph Number 15 According to unadjusted findings, measurement error (as a percentage of average MVPA) showed no evidence of change between baseline and 6-month follow-up within the control group (p = .164), but it did change from approximately 19%

lower to 10% higher within the intervention group (P = .001; Table 2). While there was no evidence of a difference between groups in mean measurement error at baseline (P = 0.157), at follow-up the mean measurement error differed significantly between groups (P = 0.010). The difference was such that at follow-up control group participants reported significantly less activity than was measured by accelerometer (-24.2%, 95% CI: -43.7, -4.7), but intervention participants did not do so (10.3, 95% CI: -6.7, 27.3). Coinciding with the differences in mean measurement error, at baseline there was no evidence of a difference between groups in the variability in measurement error (as percentage of average values; P = .497), while at follow-up, the variability in measurement error was significantly less for the intervention than the control group (P = .037). ANCOVA models showed a significant intervention effect for measurement error. At follow-up, adjusted for baseline, the intervention and control groups differed significantly in measurement error (mean difference for intervention-control = 28.0% of average MPVA, 95% CI: 3.3, 52.7, P = 0.026). As with the unadjusted results, the adjusted means showed the self-report to be significantly lower than accelerometer measures in the control group (-21.0% of average values, 95% CI: -38.2, -3.9), but not in the intervention group (6.9 % of average values, 95% CI: -10.7, 24.6).

INSERT TABLE 2 ABOUT HERE

Sensitivity Analyses

Paragraph Number 16 Results of the sensitivity analyses (Table 2) showed that the definition of accelerometer MVPA had a dramatic impact on whether self-report tended to under- or over-estimate relative to accelerometer and some impact on study conclusions regarding whether measurement error varied by group, over time, or was impacted by the intervention. The least discrepancy between self-report and accelerometer was seen in the main analysis. Self-report consistently significantly underestimated relative to the accelerometer when using the very low cutpoint for MVPA, by an unrealistic amount (i.e. up

to 7 times the average amount of MVPA). Self-report consistently significantly overestimated relative to accelerometer when using the highest cutpoint for MVPA. Significant changes over time in the intervention group and significant differences between groups at follow-up (unadjusted) were replicated with both lower and higher MVPA cutpoints, however significant control group changes in measurement error were seen with high and low cutpoints (but not the main analysis). The intervention effect on measurement error was significant in the main analysis and using the low cutpoint ($P < .001$), but not using the high cutpoint for MVPA ($P = 0.083$).

Intervention Effects for Physical Activity by Self-report and Accelerometer

Paragraph Number 17 Adjusted for baseline values, significant mean changes in MVPA (minutes/week) within the intervention group were slightly less according to the accelerometer (44.4, 95% CI: 25.0, 63.9) than self-report (72.3, 95% CI: 42.0, 102.6) (see Table, SDC 3, effect of the intervention on MVPA measured by the AAS and GT1M accelerometer). However, the non-significant changes within the control group tended to be less according to accelerometer (1.5, 95% CI: -17.4, 20.4) than by self-report (19.6, 95% CI: -10.1, 49.2), resulting in intervention effects that appeared only slightly smaller for accelerometer-measured MVPA (42.9, 95% CI: 15.5, 70.1, $P = .002$) than for self-reported MVPA (52.7, 95% CI: 10.4, 95.1, $P = .015$).

DISCUSSION

Paragraph Number 18 Self-report has limited agreement with accelerometer physical activity (18). For trial results to be unbiased in spite of measurement error requires that the error remain constant over the study duration and be equal across treatment groups. This study observed a significant intervention effect on measurement error (of self-report relative to accelerometer) in adults, and provided some evidence that measurement error can vary

significantly over the course of an intervention and is not equal for intervention and control groups at follow-up. The caution that measurement error relative to accelerometer may not be constant over time and non-differential by group is likely to apply across the vast majority of cutpoints used to assess MVPA in the literature, as both very low and very high cutpoints for MVPA replicated the changes over time within the intervention group and the group differences at follow-up (unadjusted), and the observed intervention effect on measurement error significant both in the main analysis and with the low MVPA cutpoint.

Paragraph Number 19 This trial's findings add to the few prior studies that have reported on potential changes between pre- and post- intervention in the correlation of self-report with device-based MVPA measures in adults (14, 20, 22). Two studies have noted a weakening in correlation over the course of the intervention, within both treatment groups combined (14, 20). Another has observed a strengthening in correlation within the control group and no change in correlation for the intervention group (22). However, correlations are not sensitive to shifts in systematic error. The studies suggest random error may vary over time during interventions, with the present study providing further evidence suggestive that systematic error (mean differences) also cannot safely be assumed to remain constant during intervention.

Mechanisms for Variation in Physical Activity Measurement Error

Paragraph Number 20 This study, and others comparing self-report MVPA with accelerometer (2, 7, 21), have noted both systematic error (mean differences) and random error (limits of agreement) increase with average MVPA (average of self-report and accelerometer). This may contribute to the patterning of measurement error over the intervention, which was largely consistent with the patterning of physical activity, i.e., increases in the intervention group (but not in controls) and differences between groups at follow-up (but not baseline). Observations from the study regarding random error (seen in the

limits of agreement and the variability in measurement error) were consistent with the idea that intervention participants may experience improved reporting accuracy after exposure to intervention, which included self-monitoring and education on what constitutes “moderate” and “vigorous” physical activity.

Paragraph Number 21 Issues with self-report measurement, such as misinterpretation of the survey questions, inaccurate recall, and/or social desirability bias and issues with using accelerometers in free-living populations could both contribute to the measurement error observed in this study. The mean differences, although substantial, may or may not reflect under- or over-reporting by participants, as the direction of the mean differences was very sensitive to the choice of accelerometer data treatment and therefore could just as easily represent over- or under- detection of MVPA by the accelerometer. Despite their regular use as referent assessment methods in validity studies, accelerometers are not a gold standard for physical activity measurement. The proportional increase in measurement error with increasing amount of physical activity could arise from errors in self report, if, for example, it is easier for inactive participants to recall and report infrequent, discrete bouts of physical activity than it is for active participants to recall physical activity that is completed often, and which may be variable in terms of duration, frequency and domain. The approach of applying cutpoints to define MVPA from accelerometer data could also explain the proportionality. Cut points are derived from equations that use accelerometer counts to predict energy expenditure (13), and do so very accurately for treadmill-based activities but poorly for activities of daily living (13). The imperfect sensitivity and specificity of any cutpoint for detecting each true minute of MPVA will naturally lead to an increase in the total amount of error (in minutes) as the prevalence of true minutes of MVPA increases. Type of activity may also be relevant as the most active participants may also be the most likely to engage in activities that are not well captured by accelerometers (such as cycling and swimming).

Implications for Physical Activity Interventions

Paragraph Number 22 The evidence supporting that measurement error may be affected by intervention, may vary over time and may be differential by treatment allocation has implications for comparing results from trials that use self-report exclusively with trials that use accelerometers. In the LWWD trial, physical activity changes over time within groups and intervention effects were slightly larger when measured by self-report than by accelerometer, with an additional mean change of 18.1 minutes/week for controls, 27.9 minutes/week for intervention participants and an additional 9.8 minutes/day for the intervention effect. Both measures led to the same conclusions regarding statistical significance, however, the discrepancy in effect size was sufficient to classify change as clinically relevant in the intervention group by self-report and not by accelerometer, using the LWWD a priori $\geq$ 60 minutes/week definition. The only other trial to compare intervention effects obtained by self-report (7-day physical activity recall) with accelerometer (the RT3 triaxial monitor) for the same sample of study participants (20) similarly found no difference to the statistical conclusions regarding intervention effects (i.e., both not significant) but tended to see intervention effects more in favour of the intervention group and larger changes within both intervention and control groups by self-report measures than by accelerometer measures. The results of physical activity trials may appear slightly more promising using self-report than using accelerometers, although with bias potentially arising from either measurement tool, it remains unknown whether results are exaggerated by self-report or understated by accelerometers.

Limitations

Paragraph Number 23 It should be noted that the trial was not designed and powered to detect intervention effects on measurement error and resulted in effects with wide 95% confidence intervals. Further examination in larger trials may reveal more precise evidence

regarding the extent to which measurement error is affected by intervention. The accelerometer monitoring period was scheduled without consideration of the AAS recall period; this is unlikely to affect conclusions regarding intervention effects on measurement error, as the scheduling was consistent over time and for both groups, but may have weakened the overall extent of agreement seen in this study. Importantly, as both intervention and control participants knew their activity was being monitored objectively, some biases in reporting may have been minimized in this study (e.g., such as those evidenced by sizeable self-reported improvements in unmonitored control groups). Thus, the extent of discrepancy of self-report against accelerometer in this study may be an understatement of the potential discrepancy between findings from interventions that use self-report (without objective monitoring) versus interventions that use accelerometers.

Conclusion

Paragraph Number 24 This trial showed evidence that measurement error of self-report (Active Australia Survey) relative to accelerometer was substantial and was impacted by intervention, either directly or indirectly through impact of intervention on amount of physical activity. As a result, intervention trial outcomes were estimated as slightly stronger by self-report than by accelerometer. With errors in either or both assessment tools potentially contributing to these findings, it would be prudent for interventions to measure physical activity with both self-report and device-based measures.

ACKNOWLEDGEMENTS

This study was supported by grants from the Australian National Health and Medical Research Council (NHMRC) and the Diabetes Australia Research Trust. Authors were supported by Queensland Health core infrastructure funding (Winkler); a National Heart Foundation scholarship (Waters); and, NHMRC Fellowships (Eakin, Owen and Reeves). No

other sources of funding or conflicts of interest were declared by the authors. We wish to thank the patients, general practitioners and practice staff of the Greater Metro South Brisbane Medicare Local (Logan, Australia) and Diabetes Australia Queensland. We would also like to thank project staff: Kym Spathonis, Charlotte Brakenridge, Erin Robson, Amy Chatwin, Jennifer Job, Fiona Heys, Natalie Doyle, Jodie Jetann, Ellen Baker and Charani Kiriwandeniya. The results of the present study do not constitute endorsement by the American College of Sports Medicine.

REFERENCES

1. Altman DG, Bland JM. Treatment allocation by minimisation. *BMJ*. 2005;330:843.

2. Anderson CB, Hagstromer M, Yngve A. Validation of the PDPAR as an adolescent diary: effect of accelerometer cut points. *Med Sci Sports Exerc*. 2005;37(7):1224-30.

3. Australian Government Department of Health and Ageing. *An Active Way to Better Health: National Physical Activity Guidelines for Australians*. Canberra, Australia: Australian Government Department of Health and Ageing; 1999. 2 p. Available from: Australian Government Department of Health and Ageing.

4. Australian Institute of Health and Welfare. *The Active Australia Survey: A Guide and Manual for the Implementation, Analysis and Reporting*. Canberra, Australia: Australian Institute of Health and Welfare; 2003. 55 p. Available from: Australian Institute of Health and Welfare.

5. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307-10.

6. Brown R, Richmond S. An update on the analysis of agreement for orthodontic indices. *Eur J Orthod*. 2005;27(3):286-91.

7. Brown W, Bauman A, Chey T, Trost S, Mummery K. Comparison of surveys used to measure physical activity. *Aust N Z J Public Health*. 2004;28(2):128-34.

8. Brown WJ, Burton NW, Marshall AL, Miller YD. Reliability and validity of a modified self-administered version of the Active Australia physical activity survey in a sample of mid-age women. *Aust N Z J Public Health*. 2008;32(6):535-41.

9. Brown WJ, Trost SG, Bauman A, Mummery K, Owen N. Test-retest reliability of four physical activity measures used in population surveys. *J Sci Med Sport*. 2004;7(2):205-15.

10. Eakin E, Reeves M, Marshall AL, et al. Living Well with Diabetes: a randomized controlled trial of a telephone-delivered intervention for maintenance of weight loss, physical activity and glycaemic control in adults with type 2 diabetes. *BMC Public Health*. 2010;10(1):452.

11. Ferrari P, Friedenreich C, Matthews CE. The role of measurement error in estimating levels of physical activity. *Am J Epidemiol*. 2007;166(7):832-40.

12. Freedson PS, Melanson E, Sirard J. Calibration of the computer science and applications, Inc. accelerometer. *Med Sci Sports Exerc*. 1998;30(5):777-81.

13. Matthew CE. Calibration of accelerometer output for adults. *Med Sci Sports Exerc*. 2005;37(Suppl 11):S512-22.

14. Napolitano M, Borradaile K, Lewis B, et al. Accelerometer use in a physical activity intervention trial. *Contemp Clin Trials*. 2010;31(6):514-23.

15. National Cancer Institute - SAS Programs for Analyzing NHANES 2003-2004 Accelerometer Data [Internet]. Bethesda, MD: National Cancer Institute - U.S. National Institutes of Health; [cited 2012 Jun 4]. Available from: http://riskfactor.cancer.gov/tools/nhanes_pam/.

16. Pettee GK, McClain JJ, Lee CD, et al. Evaluation of physical activity measures used in middle-aged women. *Med Sci Sports Exerc*. 2009;41(7):1403-12.

17. Prince SA, Adamo KB, Hamel ME, Hardt J, Gorber SC, 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.

18. Sallis JF, Saelens B. Assessment of physical activity by self-report: Status, limitations and future directions. *Res Q Exerc Sport*. 2000;71(2):1-14.

19. Scott NW, McPherson GC, Ramsay CR, Campbell MK. The method of minimization for allocation to clinical trials: a review. *Control Clin Trials*. 2002;23(6):662-74.

20. Sloane R, Snyder DC, Demark-Wahnefried W, Lobach D, Kraus WE. Comparing the 7-day physical activity recall with a triaxial accelerometer for measuring time in exercise. *Med Sci Sports Exerc*. 2009;41(6):1334-40.

21. Slootmaker SM, Schuit AJ, Chinapaw MJ, Seidell JC, van Mechelen W. Disagreement in physical activity assessed by accelerometer and self-report in subgroups of age, gender, education and weight status. *Int J Behav Nutr Phys Act*. 2009;6:17.

22. Taber DR, Stevens J, Murray DM, et al. The effect of a physical activity intervention on bias in self-reported activity. *Ann Epidemiol*. 2009;19(5):316-22.

23. Timperio A, Salmon J, Ball K. Evidence-based strategies to promote physical activity among children, adolescents and young adults: review and update. *J Sci Med Sport*. 2004;7(Suppl 1):20-9.

24. Timperio A, Salmon J, Rosenberg M, Bull FC. Do logbooks influence recall of physical activity in validation studies? *Med Sci Sports Exerc*. 2004;36(7):1181-6.

25. Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc*. 2007;40(1):181-8.

26. van Sluijs E, Griffin S, van Poppel M. A cross-sectional study of awareness of physical activity: associations with personal, behavioral and psychosocial factors. *Int J Behav Nutr Phys Act*. 2007;4(1):53.

27. Winkler EA, Gardiner PA, Clark BK, Matthews CE, Owen N, Healy GN. Identifying sedentary time using automated estimates of accelerometer wear time. *Br J Sports Med*. 2011;46(6):436-42.

28. World Health Organization Global Database on Body Mass Index [Internet]. Geneva, Switzerland: World Health Organization; [cited 2012 Jun 4]. Available from: http://apps.who.int/bmi/index.jsp.

FIGURE 1 - Bland Altman plots of the difference between self-reported MVPA (Active Australia Survey AAS) and accelerometer (GT1M) across average values (mean of AAS and GT1M) at baseline for a) intervention and b) control and at 6-months for c) intervention and d) control.

SUPPLEMENTAL DIGITAL CONTENT

Supplemental Digital Content 1 - Word 97-2003 document (table) - Comparison of study participants with non-participants on demographic, health, and behavioral characteristics

Supplemental Digital Content 2 - Word 97-2003 document (table) - Comparison of completers (n = 272) with those missing 6-month study outcomes (n = 30)

Supplemental Digital Content 3 - Word 97-2003 document (table) - Effect of the intervention on moderate-to-vigorous physical activity (MVPA) measured by the Active Australia Survey (AAS) and GT1M accelerometer

TABLE 1. Baseline characteristics of study participants randomized to Telephone Counseling (*n* = 151) and Usual Care (*n* = 151).

	Telephone Counseling (*n*=151)	Usual care (*n*=151)	All (*n*=302)
Age, years, mean (SD)	57.7 (8.1)	58.3 (9.0)	58.0 (8.6)
Male, *n* (%)	84 (55.6)	86 (57.0)	170 (56.3)
Body Mass Index, mean (SD)	33.1 (6.3)	33.2 (6.0)	32.3 (6.1)
Duration diabetes, years, median (25^{th}, 75^{th} percentile)	4 (2, 7)	5 (2, 10)	5 (2, 10)
Diabetes medication [a]			
Traditional OHAs, *n* (%)	114 (75.5)	119 (78.8)	233 (77.2)
Insulin, *n* (%)	23 (15.2)	20 (13.2)	43 (14.2)
New agents, *n* (%)	7 (4.6)	5 (3.3)	12 (4.0)
Other chronic conditions			
CVD related condition, *n* (%)	127 (84.1)	113 (74.8)	240 (79.5)
Musculoskeletal condition, *n* (%)	51 (33.8)	50 (33.1)	101 (33.4)
Lung condition, *n* (%)	14 (9.3)	18 (11.9)	32 (10.6)
Smoking status, *n* (%)			
Never smoker	77 (51.0)	67 (44.4)	144 (47.7)
Ex-smoker	60 (49.7)	67 (44.4)	127 (42.1)
Current smoker	14 (9.3)	17 (11.3)	31 (10.3)
Born in Australia, *n* (%)	99 (65.6)	108 (71.5)	207 (68.5)
Caucasian, *n* (%)	131 (86.8)	133 (88.1)	264 (87.4)

Employment, *n* (%)			
Full-time/Part-time or casual	97 (64.3)	93 (61.6)	190 (62.9)
Retired	40 (26.5)	42 (27.8)	82 (27.2)
Other	14 (9.3)	16 (10.6)	30 (9.9)
Income <$1000/week, *n* (%)	49 (32.5)	61 (40.4)	110 (36.4)
< High school education, *n*(%)	9 (6.0)	26 (17.2)	35 (11.6)
HbA1c, median (25^{th}, 75^{th} percentile)	7.6 (6.3, 8.5)	7.0 (6.4, 7.9)	7.1 (6.4, 8.0)
Energy intake, mean (SD)	7.1 (2.3)	6.9 (2.2)	7.0 (2.2)
Diet Quality (0-100), mean (SD)	65.6 (13.6)	65.5 (10.7)	65.6 (11.0)
Physical activity, mins/week, median (25^{th}, 75^{th} percentile)			
Self-report [b]	90.0 (20.0, 160.0)	75.0 (0.0, 200.0)	
Accelerometer [c]	93.5 (28.8, 151.9)	92.2 (39.2, 185.1)	92.7 (38.4, 180.5)

[a] OHAs = oral hypoglycaeamic medications; new agents = glucagon-like peptide-1 receptor agonists (e.g. Exenatide) or dipeptidyl peptidase-4 inhibitors (e.g. Sitagliptin)

[b] Active Australia Survey walking, moderate and vigorous activity, without doubling of the vigorous component

[c] Time spent at >=1952 counts per minute, Actigraph GT1M accelerometer

TABLE 1. Baseline characteristics of study participants randomized to Telephone Counseling (*n* = 151) and Usual Care (*n* = 151).

	Telephone Counseling (*n*=151)	Usual care (*n*=151)	All (*n*=302)
Age, years, mean (SD)	57.7 (8.1)	58.3 (9.0)	58.0 (8.6)
Male, *n* (%)	84 (55.6)	86 (57.0)	170 (56.3)
Body Mass Index, mean (SD)	33.1 (6.3)	33.2 (6.0)	32.3 (6.1)
Duration diabetes, years, median (25^{th}, 75^{th} percentile)	4 (2, 7)	5 (2, 10)	5 (2, 10)
Diabetes medication [a]			
Traditional OHAs, *n* (%)	114 (75.5)	119 (78.8)	233 (77.2)
Insulin, *n* (%)	23 (15.2)	20 (13.2)	43 (14.2)
New agents, *n* (%)	7 (4.6)	5 (3.3)	12 (4.0)
Other chronic conditions			
CVD related condition, *n* (%)	127 (84.1)	113 (74.8)	240 (79.5)
Musculoskeletal condition, *n* (%)	51 (33.8)	50 (33.1)	101 (33.4)
Lung condition, *n* (%)	14 (9.3)	18 (11.9)	32 (10.6)
Smoking status, *n* (%)			
Never smoker	77 (51.0)	67 (44.4)	144 (47.7)
Ex-smoker	60 (49.7)	67 (44.4)	127 (42.1)
Current smoker	14 (9.3)	17 (11.3)	31 (10.3)
Born in Australia, *n* (%)	99 (65.6)	108 (71.5)	207 (68.5)
Caucasian, *n* (%)	131 (86.8)	133 (88.1)	264 (87.4)
Employment, *n* (%)			
Full-time/Part-time or casual	97 (64.3)	93 (61.6)	190 (62.9)
Retired	40 (26.5)	42 (27.8)	82 (27.2)

Other	14 (9.3)	16 (10.6)	30 (9.9)
Income <$1000/week, *n* (%)	49 (32.5)	61 (40.4)	110 (36.4)
< High school education, *n*(%)	9 (6.0)	26 (17.2)	35 (11.6)
HbA1c, median (25^{th}, 75^{th} percentile)	7.6 (6.3, 8.5)	7.0 (6.4, 7.9)	7.1 (6.4, 8.0)
Energy intake, mean (SD)	7.1 (2.3)	6.9 (2.2)	7.0 (2.2)
Diet Quality (0-100), mean (SD)	65.6 (13.6)	65.5 (10.7)	65.6 (11.0)
Physical activity, mins/week, median (25^{th}, 75^{th} percentile)			
Self-report [b]	90.0 (20.0, 160.0)	75.0 (0.0, 200.0)	
Accelerometer [c]	93.5 (28.8, 151.9)	92.2 (39.2, 185.1)	92.7 (38.4, 180.5)

[a] OHAs = oral hypoglycaeamic medications; new agents = glucagon-like peptide-1 receptor agonists (e.g. Exenatide) or dipeptidyl peptidase-4 inhibitors (e.g. Sitagliptin)

[b] Active Australia Survey walking, moderate and vigorous activity, without doubling of the vigorous component

[c] Time spent at >=1952 counts per minute, Actigraph GT1M accelerometer

TABLE 2. Effect of the intervention on the difference in moderate-to-vigorous physical activity (MVPA) between the Active Australia Survey (AAS) and GT1M accelerometer as a percentage of average MVPA, using three different cutpoints for MVPA.

	Telephone Counseling (n=132)	Usual Care (n=140)	Telephone Counseling – Usual Care	
	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	*P* [a]
Freedson (≥1952 counts/minute)				
Baseline	-19.2 (-39.6, 1.2)	-40.1 (-60.5, -19.7)	20.9 (-8.1, 49.9)	0.157
6-months	10.3 (-6.7, 27.3)**	-24.2 (-43.7, -4.7)	34.5 (8.5, 60.5)	0.010[c]
6-months, adjusted for baseline[b]	6.9 (-10.7, 24.6)	-21.0 (-38.2, -3.9)	28.0 (3.3, 52.7)	0.026
Sensitivity analyses				
Low MVPA cutpoint (≥574 cpm)				
Baseline	-144.5 (-153.8, -135.1)	-149.1 (-158.0, -140.2)	4.6 (-8.4, 17.5)	0.487
6-months	-542.3 (-607.9, -476.7)***	-713.7 (-789.3, -638.0)***	171.4 (70.8, 272.0)	0.001[c]
6-months, adjusted for baseline [b]	-547.9 (-617.1, -478.8)	-708.3 (-775.5, -641.1)	160.4 (63.9, 256.9)	0.001
High MVPA cutpoint (≥2743 cpm)[d]				
Baseline	108.7 (95.6, 121.8)	95.7 (81.8, 109.6)	13.0 (-6.2, 32.2)	0.185
6-months	75.5 (57.4, 93.5)***	45.1 (22.5, 67.7)***	30.4 (1.4, 59.5)	0.040[c]

6-months, adjusted for baseline [b]	72.7 (52.4, 93.0)	47.7 (28.0, 67.4)	25.0 (-3.3, 53.3)	0.083

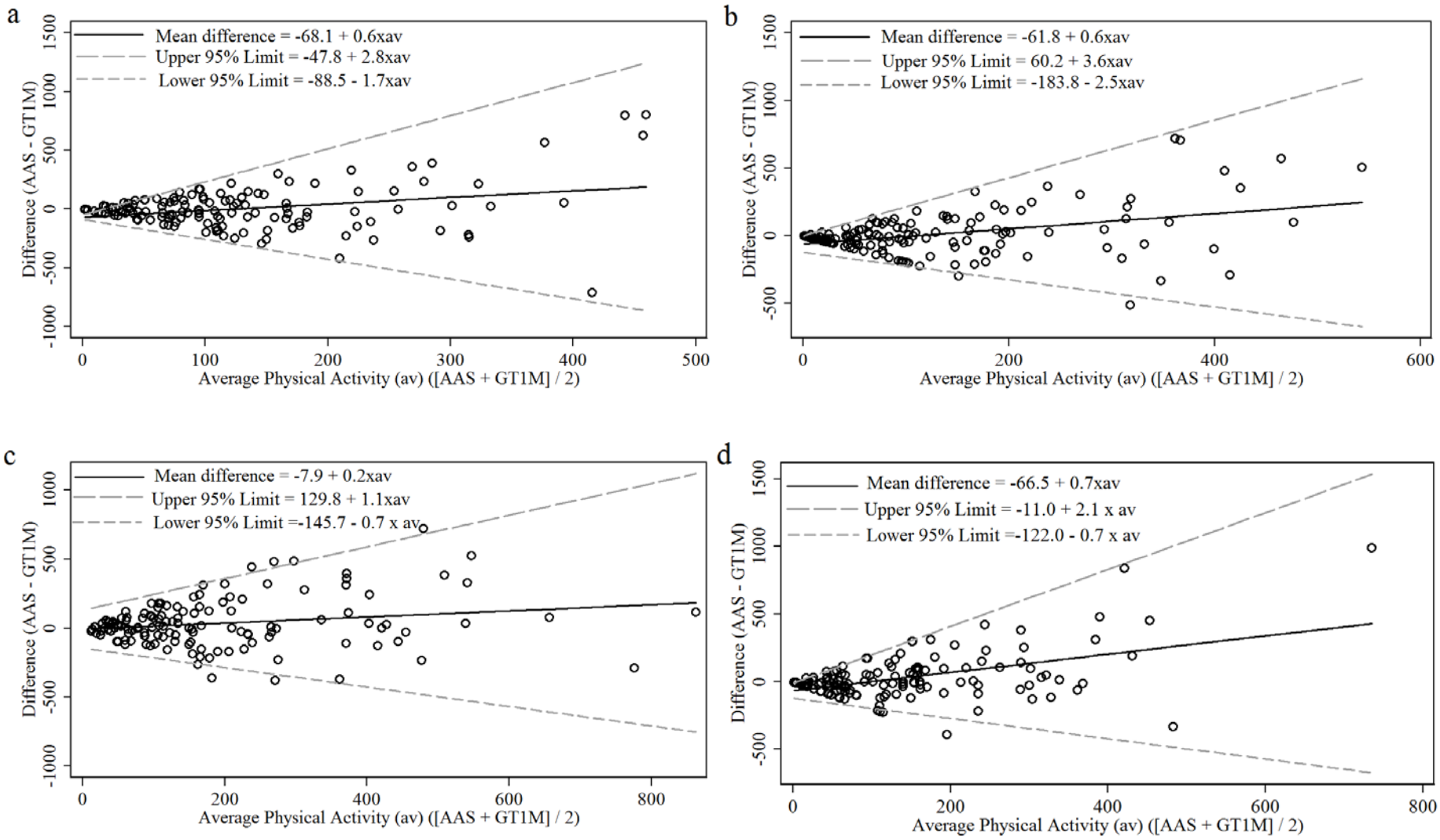
[a] *P* for difference between groups (independent samples t-test or ANCOVA)

[b] Adjusted means (95% CIs) and *P*-value from Analysis of Covariance, adjusted for baseline values, with the outcome being difference between measures as a percentage of physical activity performed at 6-months, i.e. 100*(AAS- GT1M)/(AAS+GT1M/2)

[c] Levene's test significant at $P < 0.05$; equal variance not assumed in t-test

[d] The outcome was modelled adding a small constant (0.001) to AAS and GT1M data due to values of zero average physical activity, i.e. outcome = 100*(AAS- GT1M)/([AAS+GT1M+0.002]/2)

* $P < .05$ ** $P < .01$ *** $P < .001$ for change from baseline (paired t-test)


a
Mean difference = -68.1 + 0.6xav
Upper 95% Limit = -47.8 + 2.8xav
Lower 95% Limit = -88.5 - 1.7xav
Difference (AAS - GT1M)
-1000
-500
0
500
1000
1500
0
100
200
300
400
500
Average Physical Activity (av) ([AAS + GT1M] / 2)
b
Mean difference = -61.8 + 0.6xav
Upper 95% Limit = 60.2 + 3.6xav
Lower 95% Limit = -183.8 - 2.5xav
Difference (AAS - GT1M)
-500
0
500
1000
1500
0
200
400
600
Average Physical Activity (av) ([AAS + GT1M] / 2)
c
Mean difference = -7.9 + 0.2xav
Upper 95% Limit = 129.8 + 1.1xav
Lower 95% Limit =-145.7 - 0.7 x av
Difference (AAS - GT1M)
-1000
-500
0
500
1000
0
200
400
600
800
Average Physical Activity (av) ([AAS + GT1M] / 2)
d
Mean difference = -66.5 + 0.7xav
Upper 95% Limit = -11.0 + 2.1 x av
Lower 95% Limit =-122.0 - 0.7 x av
Difference (AAS - GT1M)
-500
0
500
1000
1500
0
200
400
600
800
Average Physical Activity (av) ([AAS + GT1M] / 2)

SUPPLEMENTAL TABLE 1. Comparison of study participants with non-participants on demographic, health, and behavioral characteristics.

	Non-participants		Participants	*P* [a]
	n	value	(*n* = 302)	
Age, years, mean (SD)	111	58.4 (10.3)	58.0 (8.6)	.681
Male, *n* (%)	115	58 (50.4)	170 (56.3)	.322
Non-participant Questionnaire				
BMI (self-report), mean (SD)	64	30.6 (4.8)	32.3 (6.1)	.040
Self-report diabetes management, *n* (%)[b]				
Insulin	63	12 (19.0)	44 (14.6)	.441
Traditional OHAs	63	43 (68.3)	231 (76.5)	.200
New agents	63	2 (3.2)	7 (2.3)	.657
Lifestyle only	63	11 (17.5)	55 (18.2)	>.999
Born in Australia, *n* (%)	63	35 (55.6)	207 (68.5)	.057
Caucasian, *n* (%)	61	51 (83.6)	264 (87.4)	.411
3+ chronic conditions, *n* (%)	66	46 (69.7)	184 (60.9)	.208
Smoking status, *n* (%)	63			<.001
Never smoker		5 (7.9)	144 (47.7)	
Ex-smoker		51 (81.0)	127 (42.1)	
Current smoker		7 (11.1)	31 (10.3)	
Employment status, *n* (%)	63			.173
Full-time/Part-time/Casual		32 (50.8)	190 (62.9)	
Retired		21 (33.3)	82 (27.2)	

Other		10 (15.9)	30 (9.9)	
< High School Education, *n* (%)	63	16 (25.4)	35 (11.6)	.008
Income <$1000/week, *n* (%)	55	21 (38.2%)	110 (36.9)	.880
Married/living together, *n* (%)	63	47 (74.6)	248 (82.1)	.217
Diabetes Duration, median (25th, 75th percentile)	63	7.0 (4.0, 11.0)	5.0 (2.0, 10.0)	.005
≥5 days/week of ≥30 mins PA, *n* (%)[c]	66	13 (19.7)	57 (19.0)	.864

[a] *P* for difference between participants and non-participants by chi-square test for *n* (%), independent samples t-test for mean (Standard deviation, SD), or independent samples median test for median (25^{th}, 75^{th} percentile)

[b] OHAs = oral hypoglycaeamic medications; new agents = glucagon-like peptide-1 receptor agonists (e.g. Exenatide) or dipeptidyl peptidase-4 inhibitors (e.g. Sitagliptin)

[c] Due to missing data, *n* = 300 participants for days per week of at least 30 minutes of physical activity (PA), a single item screening question asked of most participants and in the non-participant questionnaire

SUPPLEMENTAL TABLE 2. Comparison of completers (*n* = 272) with those missing 6-month study outcomes (*n* = 30).

	Missing data (*n* = 30, 9.93%)	Completer (*n* = 272)	*P* [a]
Telephone Counseling, *n* (%)	19 (63.3)	132 (48.5)	.177
Age, years, mean (SD)	58.0 (9.2)	58.0 (8.5)	.973
Male, *n* (%)	13 (43.3)	157 (57.7)	.174
Diabetes management			
Using Insulin, *n* (%)	9 (30.0)	34 (12.5)	.023
Using traditional OHAs, *n* (%)	25 (83.3)	208 (76.5)	.496
Diabetes duration, median (25^{th}, 75^{th} percentile)	6.0 (2.8, 10.0)	4.0 (2.0, 9.8)	.172
3+ Chronic conditions, *n* (%)	20 (66.7)	164 (60.3)	.559
Smoking status, *n* (%)			.036
Never smoker	12 (40.0)	132 (48.5)	
Ex-smoker	10 (33.3)	117 (43.0)	
Current smoker	8 (26.7)	23 (8.5)	
Born in Australia, *n* (%)	20 (66.7)	187 (68.8)	.837
Caucasian, *n* (%)	28 (93.3)	236 (86.8)	.396
Income <$1000/week, *n* (%)	12 (40.0)	98 (36.0)	.692
< High school education, *n* (%)	2 (6.7)	33 (12.1)	.551
Married/living together, *n* (%)	21 (70.0)	227 (83.5)	.080
Employment, *n* (%)			.788
Full-time/Part-time/casual	20 (66.7)	170 (62.5)	
Retired	8 (26.7)	74 (27.2)	

Other	2 (6.7)	28 (10.3)	
Body Mass Index, kg/m^2, mean(SD)	33.7 (8.5)	33.1 (5.8)	.683
HbA1C, median (25th, 75th percentile)	7.6 (6.3, 8.5)	7.0 (6.4, 7.9)	.218
Energy intake, MJ, mean (SD)	6.5 (2.1)	7.0 (2.3)	.229
Diet Quality Index, 0-100, mean (SD)	65.6 (13.6)	65.5 (10.7)	.977
Physical activity, mins/week median (25th, 75th percentile)			
Self-report [b]	32.5 (0.0, 120.0)	80.0 (20.0, 180.0)	.153
Accelerometer [c]	93.5 (28.8, 151.9)	92.2 (39.2, 185.1)	.847

[a] *P* for difference between those missing data and completers by chi-square test for *n* (%), independent samples t-test for mean (Standard deviation, SD), or independent samples median test for median (25th, 75th percentile)

[b] Active Australia Survey, without doubling of the vigorous component

[c] Time spent at >=1952 counts per minute

SUPPLEMENTAL TABLE 3. Effect of the intervention on moderate-to-vigorous physical activity (MVPA) measured by the Active Australia Survey (AAS) and GT1M accelerometer.

	Baseline	6-months	Change, adjusted for baseline [a]
	Mean (SD)	Mean (SD)	Adjusted mean (95% CI)
Self-Report			
Telephone Counseling (n=135)	129.5 (157.5)	202.7 (195.7)	72.3 (42.0, 102.6)
Usual Care (n=141)	132.4 (169.4)	151.0 (179.2)	19.6 (-10.1, 49.2)
Telephone Counseling-Usual Care			52.7 (10.4, 95.1)
Accelerometer, >=1952 counts/min			
Telephone Counseling (n=133)	125.9 (116.4)	169.9 (166.0)	44.4 (25.0, 63.9)
Usual Care (n=140)	122.9 (115.6)	124.9 (110.4)	1.5 (-17.4, 20.4)
Telephone Counseling-Usual Care			42.9 (15.8, 70.1)

[a] Adjusted means (95% CIs) from Analysis of Covariance, adjusted for baseline values