Section: Brief Report

Article Title: Responsiveness to Change of Self-Report and Device-Based Physical Activity Measures in the Living Well With Diabetes Trial

Authors: Winnie Y.H. Lee, Bronwyn Kay Clark, Elisabeth Winkler, Elizabeth G. Eakin, and Marina M. Reeves

Affiliations: Cancer Prevention Research Centre, School of Population Health, University of Queensland, Brisbane, Queensland, Australia.

Running Head: Responsiveness of physical activity measures

Journal: *Journal of Physical Activity & Health*

Acceptance Date: September 3, 2014

DOI: http://dx.doi.org/10.1123/jpah.2013-0265

TITLE PAGE

Title: Responsiveness to change of self-report and device-based physical activity measures in the Living Well with Diabetes trial

Running title: Responsiveness of physical activity measures

Manuscript type: Brief Report (15 or fewer pages)

Abstract word count: 200

Manuscript word count (inclusive of all pages except the abstract and title page): 4552

ABSTRACT

Background: This study evaluated the responsiveness to change in physical activity of two self-report measures and an accelerometer in the context of a weight loss intervention trial. **Methods:** 302 participants (aged 20-75 years) with type 2 diabetes were randomised into telephone counselling (n=151) or usual care (n=151) groups. Physical activity (minutes/week) was assessed at baseline and six-months using the Active Australia Survey (AAS), the United States National Health Interview Survey (USNHIS) walking for exercise items, and accelerometer (Actigraph GT1M; ≥1952 counts/minute). Responsiveness to change was calculated as responsiveness index (RI), Cohen's d (post-scores) and Cohen's d (change-scores). **Results**: All instruments showed significant improvement in the intervention group ($p<0.001$) and no significant change for usual care ($p>0.05$). Accelerometer consistently ranked as the most responsive instrument while the least responsive was the USHNIS (responsiveness index) or AAS (Cohen's d). RIs for AAS, USNHIS and accelerometer did not differ significantly and were, respectively: 0.45 (95%CI: 0.26, 0.65); 0.38 (95%CI: 0.20, 0.56); and, 0.49 (95%CI: 0.23, 0.74). **Conclusions:** Accelerometer tended to have the highest responsiveness but differences were small and not statistically significant. Consideration of factors, such as validity, feasibility and cost, in addition to responsiveness, is important for instrument selection in future trials

Keywords: exercise, measurement, questionnaires, accelerometers, sensitivity

.

INTRODUCTION

Identifying effective interventions to increase participation in at least moderate levels of physical activity is a public health priority.[1,2] An important aspect of any physical activity intervention trial is the ability to assess change in physical activity in a valid, reliable and responsive manner.[3,4] Self-report and device-based measures are arguably the most practical ways to do so.[5] Given their convenience and low cost, self-report measures of physical activity are commonly employed,[6] especially in broad reach interventions that are often delivered by remote means, such as email, telephone, internet or mail.[7,8]

There is an extensive evidence base on the validity and reliability of self-report measures of physical activity.[9] However, there is limited evidence on their responsiveness to changes over time as a result of intervention.[9,10] Moreover, responsiveness of self-report measures has rarely been compared to device-based measures of physical activity; and when it has been examined, responsiveness has generally been assessed via the correlation between change scores, arguably not the preferred statistical approach.[11]

This study aims to examine and compare the responsiveness to change of two self-report physical activity instruments, the Active Australia Survey (AAS) and United States National Health Interview Survey (USNHIS) walking for exercise items, and accelerometer-derived physical activity.

METHODS

Study Context and Population

Data are from the Living Well with Diabetes trial, a two-arm randomised controlled trial evaluating an 18-month telephone delivered weight loss (via physical activity and dietary change) intervention compared to usual care. Interim (six-month) outcomes showed the trial achieved significant benefits for intervention over usual care in weight loss (-1.12; 95% CI: -

1.92, -0.33 % of bodyweight), mean physical activity (RR=1.30; 95% CI: 1.08, 1.57) and dietary energy intake (-0.63 MJ, 95% CI: -1.01, -0.25).[12]

Study methods have been previously described in detail.[13] In brief, participants (n=302), all adults with type 2 diabetes mellitus, aged 20-75 years, were recruited via electronic medical records search from nine general practices in Logan city, a large ethnically and socio-economically diverse community in Queensland, Australia. Participants were excluded if: i) they met or exceeded the National Physical Activity Guidelines of 30 minutes of moderate to vigorous physical activity on five or more days/week; or ii) they were not overweight or obese (BMI $\leq 25.0 kg/m^2$); and/or iii) they were currently using weight loss medications; and/or iv) they had previous or planned bariatric surgery. Ethical approval was granted from The University of Queensland Behavioural and Social Sciences Ethical Review Committee; and informed consent was obtained from participants.

Study Groups

Following baseline data collection, participants were randomised (by the minimization method[14] using the MINIM program; www.sghms.ac.uk/depts/phs/guide/randser.htm) into two groups. i) *Telephone Counselling*: Participants received a detailed workbook at the beginning of the intervention and up to 14 telephone calls over the first six months (four initial weekly calls followed by fortnightly calls) targeting physical activity, dietary intake, and weight loss consistent with management goals for type 2 diabetes mellitus.[13] Physical activity recommendations included increasing planned activity of walking or other moderate-to-vigorous activities to 30 minutes per day; increasing strength/resistance exercises to two to three sessions per week; and reducing sitting time to less than two hours per day of leisure television and computer time. Counsellors, with background in nutrition and dietetics, and with additional training in physical activity promotion and health coaching, provided targeted feedback and facilitated goal setting to promote behavioral changes and gradual weight loss during the telephone calls. ii) *Usual care*: Participants received thank you letters and a brief summary of results following each assessment throughout the duration of the study. At baseline and six-months, participants were posted standard, off-the-shelf diabetes self-management education brochures with information on diabetes education and a variety of health behaviors.

Data collection

The current study used data collected at baseline and six-months. Data were collected at home visits by registered nurses and telephone interview by research staff. All data collectors were blind to participants' group allocation. Height was measured using a portable stadiometer (Seca 214 height rod, Seca, Germany; baseline only); weight was measured without shoes or heavy clothing to the nearest 0.1 kg using standard calibrated scales (Model TI TBF 350, Tanita Inc., Tokyo, Japan). Telephone interviews collected self-report data on demographics/ selected health characteristics (age, gender, number of chronic conditions, current smoking, ethnicity, employment status, income, and educational attainment) and physical activity.

Physical activity was assessed using the Active Australia Survey (AAS), walking for exercise items from the United States National Health Interview Survey (USNHIS), and accelerometer. The AAS is an 8-item measure used to assess number of days and total duration spent walking and in moderate- and vigorous-intensity physical activities over the last seven days. The AAS has shown moderate to good test-retest reliability for measuring total physical activity (ICC= 0.59 [95% CI: 0.52, 0.65][15] and $\rho = 0.80$ [95% CI: 0.68, 0.87]).[16]Validity of the AAS as a telephone-administered survey compared to accelerometer-derived moderate and vigorous physical activity (≥1952 counts per minute) has shown correlations (ρ) between0.42 to 0.61.[16,17] As per the AAS guidelines, total minutes/week of physical activity were calculated from the sum of walking, moderate, and two × vigorous (excluding gardening) durations.[18] To reduce over-reporting, each activity was truncated at 840 minutes and total physical activity was truncated at 1680 minutes.[18] Two walking for exercise items were adapted from the USNHIS to assess the frequency and duration spent walking for exercise. The duration item was adapted to enquire about the amount of time spent walking for exercise in the past week rather than the amount of time per session.[19] The USNHIS walking for exercise items showed a correlation of r = 0.36 compared with a 1-week log of walking behavior.[20]

Objective data on physical activity were collected via GT1M accelerometers (Actigraph, LLC, Fort Walton Beach, Florida) set to record data in one minute epochs. Participants were provided with the accelerometer on an adjustable belt that was fitted firmly around the waist, positioned on the right mid-axillary line, to detect any ambulatory movements. Participants were instructed to wear the monitor at all times while awake (except during water-based activities) for a continuous period of seven days, and to record all monitor removals in a log. Non-wear times were removed, having been identified by research staff as the precise times movement began or stopped that best coincided with their self-reported wear/removal periods. For each day, moderate to vigorous physical activity (MVPA) was identified as all worn minutes ≥1952 counts.[21] Daily MVPA was averaged for valid days (10+ hours of wear) and then multiplied by 7 to yield a weekly estimate (minutes/week). To reduce reliance on cutpoints, and examine total physical activity (not just MVPA), the average per valid day of total accelerometer counts was also calculated, along with counts per minute (total counts/ wear time). Accelerometer compliance was high, with almost all participants (98%; 297/302) and 6-month completers (97%; 264/272) providing at least 4 valid days of data at baseline and 6-months, respectively. Mean (±standard deviation) daily wear time was 13.5±1.6 hours at baseline and 13.7±1.7 at 6-months.

Statistical Analyses

Data analyses were performed using SPSS version 20.0 (IBM Corporation) with statistical significance set at $p<0.05$, two-tailed. Analyses focused on participants with complete physical activity data at baseline and 6-month follow-up. The 6-month follow-up was selected for this analysis as this was the intensive phase of the intervention. Participant characteristics at baseline are described as n (%), mean (standard deviation, SD) or median (25^{th} and 75^{th} percentiles) for non-normal data. Physical activity data were skewed, but changes approximated normality.

Responsiveness

Several measures of responsiveness exist and these can yield varied results.[22] Accordingly, we used several measures of responsiveness to ensure results are robust and for the specific advantages each method has. We examined Responsiveness Index (RI; intervention mean change/ usual care standard deviation of change) with 95% Confidence Intervalsas outlined by Tuley [23] andCohen's d, firstly based on post-scores ([intervention mean − usual care mean] /pooled standard deviation), and secondly based on changes ([intervention mean change − usual care mean change] /pooled standard deviation of change), with 95% Confidence Intervals obtained using a macro EFFECT_CI[24] in SAS (SAS Institute Inc.100 SAS Campus Drive Cary, NC 27513-2414, USA). Single arm trials also routinely report paired t tests in assessing responsiveness; in view of our two-group design we report intervention effects (independent samples t-tests of change scores). RI was used to statistically compare responsiveness across instruments.[23] Cohen's d (post-scores) were reported despite their problematic use with skewed data due to their useful interpretation of effect sizes as trivial (<0.20), small (≥0.20 to <0.5), or large (≥0.80). Responsiveness was assessed for the various measures but only compared between instruments on the most comparable measure, minutes/week of MVPA.

RESULTS

Participant characteristics

Table 1 shows participants' characteristics at baseline by group. The mean (±SD) age for all participants was 58.0 (±8.6) years and 56.3% were men. Nearly all participants were either overweight (26.2%) or obese (68.2%) and were primarily Caucasian (87.4%). These characteristics are similar to the Australian population of adults with type 2 diabetes as reported for the nationally representative AusDiab study.[25]

Of the 302 participants recruited into this study, complete physical activity data were available for 272 (90.1%) participants at six-months (87.4% of telephone counselling and 92.7% of usual care). Loss to follow-up (missing data) was minimal in both groups, non-differential, and significantly associated only with use of insulin (p=0.023) and smoking status (p=0.036).[12]

Responsiveness

Table 2 shows change in MVPA measured by the AAS, the USNHIS and accelerometer and the responsiveness to change of the three instruments over the six-month telephone counselling intervention. All three instruments showed significant improvements in physical activity within the Telephone Counselling group, ranging from 45 minutes/week [accelerometer] to 85 minutes/week [AAS]). None of the instruments detected a statistically significant change within the Usual Care group, with changes being very small for accelerometer (2 minutes/week) and USNHIS (-2 minutes/week) and slightly larger for the AAS (23 minutes/week). Intervention effects were most statistically significant (p-values) by accelerometer (p=.004), followed by the USNHIS (p=.008) then AAS (p=.022) despite size of effect being largest according to AAS (62 minutes/week) and smallest by accelerometer (43 minutes/week).

Consistently across all responsiveness statistics examined, the highest responsiveness was observed for the accelerometer. However, the accelerometer did not have a statistically significantly higher RI than AAS (difference = 0.03, 95% CI: -0.11, 0.35) or USHNIS (difference = 0.12, 95% CI: -0.11, 0.35). The least responsive instrument varied depending on choice of responsiveness statistic. Relative to the USNHIS, the AAS tended to show responsiveness that was higher (but not significantly so) according to RI (difference = 0.09, 95% CI: -0.08, 0.27), or lower according to Cohen's d (change, 0.28 versus 0.33). For all three instruments, responsiveness defined by effect size could be considered 'small' (≥0.20 to < 0.50).

Responsiveness of the accelerometer was significantly (p<0.05) greater for weekly minutes of MVPA (Responsiveness Index [95% CI] = 0.45 [0.21, 0.70]) than for total counts (Responsiveness Index [95% CI] = 0.26 [0.06, 0.46]; Cohen's d (post) [95% CI] = 0.22 [-0.02, 0.46]; Cohen's d (change) [95% CI]=0.25 [0.01, 0.49]) or average counts (Responsiveness Index [95% CI] = 0.14 [-0.04, 0.31]; Cohen's d (post) [95% CI]= 0.17 [-0.07, 0.41]; Cohen's d (change) [95% CI]= 0.16 [-0.07, 0.40]).

DISCUSSION

Understanding the ability of physical activity measures to detect longitudinal changes in intervention settings is essential. This study was one of the first to examine the responsiveness of self-report and device-based measures to detect changes in physical activity. Responsiveness was only modest (i.e. 'small' effect size) for all three instruments, with none of the instruments being significantly more responsive than others. Although in part this finding could arise from the intervention only truly achieving a modest intervention effect (between 43 mins/week by accelerometer and 62 minutes/week by AAS). In theory, tools with less measurement error could still provide better responsiveness to detect modest effects.

Notably, of the three instruments evaluated, the accelerometer consistently ranked as the most responsive by all definitions and produced the most highly significant intervention effect sizes. The accelerometer had a clear advantage in terms of smaller variability (in post scores, and in change) while the self-report tools had the advantage in terms of size of intervention changes detected, leading to similar responsiveness overall. While several responsiveness statistics were used to ensure robustness in findings, importantly all relied on some definition of observed change in the intervention in defining the numerator for responsiveness. So, while results indicated no power/sample size advantage of any tool in terms of amount of change that can be detected in a particular trial, the same could not be said of power/sample size advantages to detect a particular sized estimated difference of interest (e.g., 60 minutes/week). In this case, the accelerometer would be the most efficient, but it would be harder to achieve a 60 minute/week intervention effect by accelerometer than by AAS.

Very few studies to date have examined the ability of physical activity measures to detect changes over time.[9] Typically such studies have focused on validity of instruments to measure change, usually by examining the correlation between change scores according to self-report and some criterion measure. In a population of Latino adults, Hoos et al[11] reported a non-significant correlation (ρ=0.14) between change in moderate-to-vigorous physical activity as measured by the Global Physical Activity Questionnaire and by the accelerometer after a six-month intervention. Sloane et al.[26] also reported a non-significant correlation between change in moderate to vigorous physical activity measured by a 7-day physical activity recall and accelerometer after a 12-month intervention. Similarly, changes in total physical activity measured by the International Physical Activity Questionnaire have shown non-significant correlations with changes in VO_2 max over 3 years in an observational study (ρ=0.12-0.20).[27] More recently, a study based on this LWWD trial extended these correlation findings by focusing also on systematic error.[28] We reported that the intervention appears to impact validity, in that the discrepancy between self-report and accelerometer widens following intervention, especially within the intervention group, such that larger intervention effects are observed for self-report than accelerometer. It is unclear whether intervention effects may be overstated by self-report or understated by accelerometer as self-report tools are prone to reporting biases but accelerometers also under-detect some activities.

The only study located in the literature that examined responsiveness to change of physical activity measurement instruments using the responsiveness index was conducted by our group.[10] This study reported on responsiveness to change in self-reported physical activity after a 4-month intervention similar to the one reported here, in adults with type 2 diabetes and/or hypertension. RIs were slightly lower than in the present study, but not significantly so, being 0.25 (95% CI: 0.11, 0.39) for AAS and 0.27 (95% CI: 0.13, 0.41) for USNHIS.

The limited evidence on responsiveness to date does not support that GT1M accelerometer, AAS or USNHIS provide a highly responsive measure for detecting physical activity change. As the AAS and USNHIS have already undergone development, refinement and were administered by telephone there is likely little that can be done to enhance their responsiveness, except for using multiple administrations. Multiple administrations could reduce measurement error by reducing week-to-week variability. Similarly, responsiveness of the accelerometer in theory could be improved both by reducing measurement error, with administrations over longer periods reducing day-to-day variability. As accelerometers often under-detect certain activities such as cycling[29] and are typically not worn in water-based activities such as swimming, it is possible steps could be undertaken to reduce under-detection. For example, it is possible that under-detection might be reduced by addition of self-reported activities not observed by accelerometer, and/or some as yet unknown optimal combination of type of accelerometer (brand, uniaxial/triaxial) and approach to identifying physical activity (e.g., pattern recognition, various epochs, various cutpoints, bout restrictions).

A major strength of this study was the inclusion of device-based measurement of MVPA along with self-report measures, and the ability to compare responsiveness across these measures within a single study. The study only included data from the GT1M accelerometer, and this was analysed in 60-second epochs using the vertical axis data only, with the Freedson cutpoints applied. This is a common approach with acceptable validity,[30] but results cannot be generalised to newer model triaxial accelerometers or other methods of identifying physical activity (e.g., shorter epochs, different cut-points). Relatedly, the responsiveness of the accelerometer also depends on the physical activity measure used (e.g., weekly minutes of MVPA versus total physical activity assessed by accelerometer counts), in part due to measurement properties and partly due to the relative emphasis placed on targeting moderate and vigorous versus light and sedentary activities in the intervention. Also, the participants were adults with type 2 diabetes mellitus, primarily overweight or obese, which while largely representative of the Australian adult population with type 2 diabetes mellitus,[25] would not be representative of adults more generally. Another limitation is that both the calculation and interpretation of responsiveness rely to varying degrees on the assumption of normality, which is almost never completely satisfied with physical activity data. Nevertheless, as most researchers analyse their physical activity intervention data using methods that assume a normal distribution, these responsiveness findings are still relevant.

CONCLUSION

This study provides evidence on the ability of two commonly used self-report physical activity instruments to measure change in physical activity compared with a device-based measure of physical activity. Findings showed MVPA from the AAS, walking for exercise items from the USNHIS and MVPA measures from the Actigraph GT1M accelerometer all had modest responsiveness. While the accelerometer tended to be most responsive, no instrument was significantly more responsive than the others. Other concerns, such as validity and feasibility, may therefore be more relevant than responsiveness in choosing between these instruments to detect physical activity changes. Repeat administrations (self-report) or longer administrations (accelerometer) are simple possible solutions that ought to be investigated for studies requiring more responsive physical activity instruments.

ACKNOWLEDGEMENTS

The authors would like to thank the patients, general practitioners and practice staff of the Southeast Primary Healthcare Network (Logan, Australia) for participating in the study and Diabetes Australia Queensland for their endorsement and provision of materials for the usual care group. Further acknowledgements go to the project staff for their integrity and commitment.

FUNDING SOURCES

This study was supported by a National Health and Medical Research Council (NHMRC) project grant # 511000. The trial is registered with the Australian Clinical Trials Registry #ACTRN12608000203358.

REFERENCES

1. Haskell WL, Lee I-M, Pate RR, et al. Physical activity and public health. Updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Circulation.* 2007;116(9):1081-1093. doi: 10.1161/CIRCULATIONAHA.107.185649
2. World Health Organisation. Global recommendations on physical activity for health. Switzerland: WHO Press; 2010.
3. Hagströmer M, Ainsworth BE, Kwak L, Bowles HR. A checklist for evaluating the methodological quality of validation studies on self-report instruments for physical activity and sedentary behavior. *J Phys Act Health.* 2012;9 Suppl 1:S29-36. URL: http://www.ncbi.nlm.nih.gov/pubmed/22287445
4. Pettee Gabriel KK, Morrow JRJ, Woolsey AL. Framework for physical activity as a complex and multidimensional behavior. *J Phys Act Health.* 2012;9 Suppl 1:S11-18. URL: <Go to ISI>://WOS:000299764400003
5. Tudor-Locke CE, Myers AM. Challenges and opportunities for measuring physical activity in sedentary adults. *Sports Med.* 2001;31(2):91-100. URL: http://www.ncbi.nlm.nih.gov/pubmed/11227981
6. Washburn RA, Heath GW, Jackson AW. Reliability, validity, and methodological issues concerning large-scale surveillance of physical activity. *Res Q Exerc Sport.* 2000;71(2):104. URL: http://proquest.umi.com/pqdweb?did=56329794&Fmt=7&clientId=20806&RQT=309&VName=PQD

7. Heath GW, Parra DC, Sarmiento OL, et al. Evidence-based intervention in physical activity: lessons from around the world. *Lancet.* 2012;380(9838):272-281. doi: 10.1016/S0140-6736(12)60816-2

8. Marshall AL, Owen N, Bauman AE. Mediated approaches for influencing physical activity: update of the evidence on mass media, print, telephone and website delivery of interventions. *J Sci Med Sport.* 2004;7(1 Suppl):74-80. URL: http://www.ncbi.nlm.nih.gov/pubmed/15214605

9. van Poppel MN, Chinapaw MJ, Mokkink LB, van Mechelen W, Terwee CB. Physical activity questionnaires for adults: a systematic review of measurement properties. *Sports Med.* 2010;40(7):565-600. doi: 10.2165/11531930-000000000-00000

10. Reeves MM, Marshall AL, Owen N, Winkler E, Eakin EG. Measuring physical activity change in broad-reach intervention trials. *J Phys Act Health.* 2010;7(2):194-202. URL: <Go to ISI>://WOS:000299764400003

11. Hoos T, Espinoza N, Marshall SJ, Arredondo EM. Validity of the Global Physical Activity Questionnaire (GPAQ) in adult Latinas. *J Phys Act Health.* 2012;9(5):698-705. URL: http://www.ncbi.nlm.nih.gov/pubmed/22733873

12. Eakin E, Reeves M, Winkler E, et al. Six-month outcomes from Living Well with Diabetes: A randomized trial of a telephone delivered weight loss and physical activity intervention to improve glycemic control. *Ann Behav Med.* 2013 (Accepted for publication). doi: 10.1007/s12160-013-9498-2

13. Eakin EG, Reeves MM, 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. doi: 10.1186/1471-2458-10-452

14. Altman DG, Bland JM. Treatment allocation by minimisation. *BMJ.* 2005;330:843. doi: http://dx.doi.org/10.1136/bmj.330.7495.843

15. 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):205-215. doi: 10.1016/s1440-2440(04)80010-0

16. Fjeldsoe B, Winkler E, Marshall A, Eakin E, Reeves M. Active adults recall their physical activity differently to less active adults: test-retest reliability and validity of a physical activity survey. *Health Promot J Austr.* 2013;24(1):26-31. doi: 10.1071/HE12912

17. 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):1181-1186. doi: 10.1249/01.Mss.0000132268.74992.D8

18. Australian Institute of Health and Welfare. *The Active Australia Survey: a guide and manual for implementation, analysis and reporting.* Canberra: AIHW;2003.

19. Hovell MF, Hofstetter CR, Sallis JF, Rauh MJ, Barrington E. Correlates of change in walking for exercise: an exploratory analysis. *Res Q Exerc Sport.* 1992;63(4):425-434. URL: http://www.ncbi.nlm.nih.gov/pubmed/1439168

20. Johnson MF, Sallis JF, Hovell MF. Self-report assessment of walking: Effects of aided recall instructions and item order. *Meas Phys Educ Exerc Sci.* 2000;4(3):141-155. doi: 10.1207/S15327841Mpee0403_1

21. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. 1998;30(5):777-781. URL: file:///U:/Paper%20PDFs/Freedson%201998%20MSSE.htm

22. Stratford PW, Riddle DL. Assessing sensitivity to change: choosing the appropriate change coefficient. *Health Qual Life Outcomes.* 2005;3:23. doi: 10.1186/1477-7525-3-23

23. Tuley MR, Mulrow CD, McMahan CA. Estimating and testing an index of responsiveness and the relationship of the index to power. *J Clin Epidemiol.* 1991;44(4–5):417-421. doi: 10.1016/0895-4356(91)90080-s

24. Hess MR, Kromrey JD. EFFECT_CI: A SAS macro for constructing confidence intervals around standardized mean differences. . Proceedings of the Southeast SAS Users Group; 2003.

25. Williams ED, Tapp RJ, Magliano DJ, Shaw JE, Zimmet PZ, Oldenburg BF. Health behaviours, socioeconomic status and diabetes incidence: the Australian Diabetes Obesity and Lifestyle Study (AusDiab). *Diabetologia.* 2010;53(12):2538-2545. doi: 10.1007/s00125-010-1888-4

26. Sloane R, Snyder DC, Demark-Wahnefried W, Lobach D, WE. K. Comparing the 7-day physical activity recall with a triaxial accelerometer for measuring time in exercise. *Med Sci Sports Exerc.* 2009;41(6):1334-1340. doi: 10.1249/MSS.0b013e3181984fa8

27. Graff-Iversen S, Anderssen S, Holme I, Jenum A, Raastad T. An adapted version of the long International Physical Activity Questionnaire (IPAQ-L): construct validity in a low-income, multiethnic population study from Oslo, Norway. *Int J Behav Nutr Phys Act.* 2007;4(1):13. doi: 10.1186/1479-5868-4-13

28. Winkler E, Waters L, Eakin E, Fjeldsoe B, Owen N, Reeves M. Is measurement error altered by participation in a physical activity intervention? *Med Sci Sports Exerc.* 2013;45(5):1004-1011. doi: 10.1249/MSS.0b013e31827ccf7d

29. Wetten AA, Batterham M, Tan SY, Tapsell L. Relative validity of three accelerometer models for estimating energy expenditure during light activity. *J Phys Act Health.* 2013. URL: http://www.ncbi.nlm.nih.gov/pubmed/23417054

30. Lyden K, Kozey SL, Staudenmeyer JW, Freedson PS. A comprehensive evaluation of commonly used accelerometer energy expenditure and MET prediction equations. *Eur J Appl Physiol.* 2011;111(2):187-201. doi: 10.1007/s00421-010-1639-8

Table 1 Baseline characteristics of study participants randomized to usual care (*n*=151) and telephone counselling (*n*=151) groups

	Usual Care (*n*=151)	**Telephone Counselling (*n*=151)**	**All (*n*=302)**
Age, years, *mean (SD)*	58.3 (9.0)	57.7 (8.1)	58.0 (8.6)
Gender- Male, *n (%)*	86 (57.0)	84 (55.6)	170 (56.3)
Body Mass Index, kg/m^2, *mean (SD)*	33.2 (6.0)	33.1 (6.3)	32.3 (6.1)
Number of chronic conditions			
0, *n (%)*	21 (13.9)	13 (8.6)	34 (11.3)
1-2, *n (%)*	89 (58.9)	89 (58.9)	178 (58.9)
3-4, *n (%)*	34 (22.5)	42 (27.8)	76 (25.2)
≥ 5, *n (%)*	7 (4.7)	7 (4.7)	14 (4.7)
Number of current smokers, *n (%)*	17 (11.3)	14 (9.3)	31 (10.3)
Ethnicity- Caucasian, *n (%)*	133 (88.1)	131 (86.8)	264 (87.4)
Employment, *n (%)*			
Full-/Part-time or casual	93 (61.6)	97 (64.3)	190 (62.9)
Retired	42 (27.8)	40 (26.5)	82 (27.2)
Other	16 (10.6)	14 (9.3)	30 (9.9)
Income <$1000/week, *n (%)*	61 (40.4)	49 (32.5)	110 (36.4)
< High school education, *n (%)*	26 (17.2)	9 (6.0)	35 (11.6)
AAS [a]			
Total walking, moderate, and vigorous PA (minutes/week)	80 (0,200)	90 (20,170)	90 (20,180)
USNHIS [a]			
Walking for exercise (minutes/week)	15 (0,105)	30 (0,90)	20 (0,100)
Accelerometer [a]			
Average moderate and vigorous PA, time spent at ≥1952 cpm (minutes/week)	88 (35,182)	95 (42,178)	93 (38,181)

Abbreviations: AAS, Active Australia Survey; USNHIS, United States National Health Interview Survey; PA, physical activity.

[a] Data are median (inter-quartile range).

Table 2 Responsiveness to change over a six-month telephone counselling intervention (Living Well with Diabetes) of the Australia Active Survey (AAS, total physical activity) , the United States National Health Interview Survey (USNHIS, walking for exercise) and GT1M accelerometer (moderate-to-vigorous physical activity, ≥1952 cpm)

	GT1M	AAS	USNHIS
Physical activity mins/week, Mean±SD			
Telephone Counselling (n=132)			
Baseline	127 ± 117	143 ± 181	72 ± 118
6-months	171 ±166	228 ± 231	121 ± 150
Change	45 ± 137*	85 ± 241*	49 ± 164*
Usual Care (n=140)			
Baseline	125 ± 110	142 ± 174	80 ± 133
6-months	123 ±116	165 ± 193	78 ± 132
Change	2 ± 98	23 ± 199	-2 ± 147
Intervention effect (95% CI), p [a]	43 (14,71), p=.004	62 (9,115), p=.022	51(13,88), p=.008
Responsiveness (95% CI)			
Responsiveness Index	0.45 (0.21, 0.70)	0.43 (0.21, 0.64)	0.33 (0.14, 0.53)
Cohen's d (post)	0.33 (0.09, 0.57)	0.30 (0.06, 0.54)	0.30 (0.06, 0.54)
Cohen's d (change)	0.36 (0.12, 0.60)	0.28 (0.04, 0.52)	0.33(0.09, 0.56)
Difference in Responsiveness Index (95% CI)			
vs AAS	0.03 (-0.11, 0.35)	-	-
vs USNHIS	0.12 (-0.11, 0.35)[b]	0.09 (-0.08, 0.27)[b]	-

* paired t-test significant ($p<0.05$) change between baseline and six month follow-up.

[a] Intervention change – control change (independent samples t-test)

[b] assumption satisfied such that higher responsiveness denotes more power (ratio of responsiveness indices greater than ratio of variances)

Supplemental Table 1: Comparison of baseline characteristics of completers (n=255) and those missing six-month study outcomes (n=47)

	Missing data (n=47)	Completer (n=255)	p [a]
Age, years, *mean (SD)*	58.1 (8.3)	58.0 (8.6)	.900
Gender- Male*, n (%)*	24 (51.5)	146 (57.3)	.432
Body Mass Index, kg/m^2, *mean (SD)*	34.8 (7.8)	32.8 (5.8)	.037
Number of chronic conditions, *n (%)*			.064
0, *n (%)*	8 (17.0)	26 (10.2)	
1-2, *n (%)*	20 (42.6)	158 (62.0)	
3-4, *n (%)*	14 (29.8)	62 (24.3)	
≥ 5, *n (%)*	6 (12.8)	9 (3.5)	
Number of current smokers, *n (%)*	8 (17.0)	19 (7.5)	.007
Ethnicity- Caucasian*, n (%)*	43 (91.5)	221(86.7)	.664
Employment*, n (%)*			.929
Full-time/ Part-time/casual	31 (66.0)	159 (62.4)	
Retired	11(23.4)	71(27.8)	
Other	5 (10.6)	25 (9.8)	
Income <$1000/week*, n (%)*	28 (59.6)	137 (53.7)	.701
< High school education*, n (%)*	22 (46.8)	89 (34.9)	.576

[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)