# Physical activity and longevity: how to move closer to causal inference

Kaitlin H Wade, Rebecca C Richmond, George Davey Smith

Kujala provides an insightful review contesting epidemiological findings that increased physical activity (PA) lengthens the life span,[1] arguing that intervention (randomised controlled trial (RCT) and experimental) studies do not support PA causing a reduced risk of death and highlighting several limitations in previous observational studies that may have led to spurious conclusions.

The review coincides with the publication of findings from the large-scale Prospective Urban Rural Epidemiologic (PURE) study (n=130 843), which identified a graded lower rate of mortality among those individuals achieving moderate and high levels of PA compared with those with low PA (HR 0.80; 95% CI 0.74 to 0.87 and 0.65; 95% CI 0.60 to 0.70; P for trend <0.0001).[2] While this study is undeniably an impressive endeavour, collecting prospective data on participants from 17 countries, the conclusion to support increased forms of PA levels for all individuals (irrespective of age, gender or country of origin) has major public health implications. The findings are, as so often, qualified by the study, being unable to fully assert a causal (rather than correlational) role for PA levels in reducing mortality.

Kujala emphasises how epidemiological study designs are vulnerable to limitations that may skew or distort observational associations and impede the distinction between correlation and causation. Such distortions of observed relationships may arise due to confounding by measured/ unmeasured lifestyle, behavioural and biological factors (such as higher fitness, lower body mass index (BMI), genetic variation and socioeconomic factors) correlated with both the exposure (here, PA) and outcome (here, longevity). If not appropriately accounted for, confounding factors make the ascertainment of underlying causal mechanisms and pathways exceptionally complex. Such was illustrated by Jerry Morris' London busmen study revisited by Kujala, where

confounding by baseline adiposity biased findings that bus conductors had lower risk of coronary heart disease than their less-active driver counterparts (although this issue was acknowledged by Morris who performed analysis stratified by the busmen's uniform size to account for potential confounding by adiposity).[3]

The possibility of reverse causation (whereby the 'outcome' is responsible for variation in the 'exposure', rather than the direction of interrogation) may also lead to misinterpretation of observed associations. For example, the notion that reducing PA increases the risk of becoming overweight/obese is as plausible as the reverse, where being overweight/obese renders PA difficult.[4] Studies of older adults or those with many comorbidities are particularly vulnerable to reverse causation. For example, in reference to Kujala's 'healthy exerciser bias', aged individuals who are healthy enough to participate in PA due to a lack of chronic illness will seemingly have a reduced risk of death compared with their less-fit peers. Furthermore, comparing estimates of risk for physically demanding versus sedentary occupations may suffer reverse causation, particularly when high fitness and good health are criteria for recruitment into such physically demanding occupations.

Related to this, in the setting of evaluating potential causes of mortality, both selection and survival biases,[5] which influence participation rates in epidemiological studies, can also lead to distortion of associations among respondents. In these cases, the population under study (and therefore the observed associations) may differ from the population not selected or who were unable/unwilling to participate (due to morbidity or lack of interest in surveys relating to health).[6]

Kujala also highlights the limitation of measurement error, which can bias estimates within epidemiological studies, particularly those relying on self-report or questionnaire-based information. Recent developments have highlighted the trade-off between sample size and measurement precision in obtaining adequate statistical power with minimum measurement error. For example, measuring maximal oxygen consumption in a formal

fitness test, $VO_2max$, in a smaller sample rather than self-reported, retrospective PA in a large sample may provide a more precise predictor of mortality.[7] Furthermore, inadequate measurement, limited knowledge or poor adjustment for confounding variables, such as smoking status in the setting of physical activity and mortality, can severely bias observed associations.

As presented by Kujala, RCTs, the gold standard in epidemiology for inferring causality, have failed to provide conclusive evidence in this context (eg, Lifestyle Interventions and Independence for Elders,[8] Look Action for Health in Diabetes,[9] Heart Failure: A Controlled Trial Investigating Outcomes of Exercise Training[10] and other large-scale meta-analyses).[11 12] In the absence of long-term trials, the focus moves to other approaches for strengthening causal inference. Some such methods are discussed by Kujala and are outlined in table 1.

One approach acknowledged is the comparison of associations between recreational leisure time and obligatory occupational PA, the latter of which has not consistently been associated with a reduced risk of death.[13] Physiologically, there are no compelling reasons why recreational and occupational PA should have systematically different effects on mortality and so, if activity were truly causal, effect sizes should be similar between these two contexts. One explanation for this potential discordance is confounding by socioeconomic position. For example, earlier studies of the association between occupation and PA, at a time when there may have been a positive social class gradient for cardiovascular disease (CVD), tended to show that doing more occupational PA was related to lower CVD.[14] However, with a change in social class gradient over time, more recent studies have typically failed to demonstrate consistent and protective effects of occupational PA.[13]

While the recent PURE study found that occupational PA was protective against mortality risk across countries at different economic levels, it is important to highlight that definitions of occupational PA included travel to work, which may be strongly influenced by health-related selection.[15] Interestingly, the PURE study does not seem to be as supportive for the role of recreational activity on reducing mortality risk, where differences in underlying confounding structures between varying income countries investigated may explain the heterogeneity in effects observed.

Medical Research Council Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

**Correspondence to** Dr Kaitlin H Wade, Integrative Epidemiology Unit, University of Bristol, Bristol BS8 2BN, UK; Kaitlin.Wade@bristol.ac.uk

**Table 1**  Methods for strengthening causal inference in physical activity epidemiology studies

| Method | Example | Strength | Limitation |
|---|---|---|---|
| Lifecourse approach | Use of prospective studies to investigate PA levels at different ages and how they might differently affect lifespan | Useful for assessing temporal associations; ability to adjust for the respective outcome measures at baseline (where possible) makes it possible to disentangle prospective associations from tracking effects | Logistically demanding as it requires repeat assessments; residual confounding; measurement error in exposure, outcome and covariables; selection bias |
| Cross-context comparison | Comparison of associations between voluntary leisure-time PA and compulsory occupational PA; PA across different cultures or dissimilar countries | Exploring residual confounding; reliable findings if estimates are similar across different contexts (where the confounding structure in these settings is likely to differ) | Assumptions about different confounding structures may not be correct; variables in different studies might be measured with varying accuracy and generalisability |
| Sibling comparison | MZ or DZ twin comparisons among siblings discordant for PA | Using MZ best controls for familial background and genetic confounding, compared with DZ (or siblings), where 50% of genetic information is shared | Assumes a stable family environment; confounding by factors not perfectly shared by siblings; reverse causation still possible |
| Mendelian randomisation | The use of genetic variants associated with exercise and fitness, incorporated into a Mendelian randomisation analysis, whereby genotype serves as an instrumental variable for PA | Genetic instruments are not subject to confounding from environmental or lifestyle factors, are not influenced by the outcome, do not change over time and are measured with high accuracy | Low power; lack of instruments; pleiotropy and linkage disequilibrium; population stratification; canalisation |
| Objective measures of the exposure and biomarkers | The use of objective measures of PA (eg, accelerometry data), fitness (eg, $VO_2$ max) or the incorporation of other biomarkers (eg, DNA methylation) | More precise measurement of underlying risk factor reduces measurement error in documented PA and problems with regression dilution bias; biomarkers may serve as surrogate endpoints in PA trials and longitudinal studies without long-term follow-up | Does not evade problems of confounding or reverse causation |

DZ, dizygotic; MZ, monozygotic; PA, physical activity.

A further causal inference approach not directly considered by Kujala is that of the negative control situation, which involves using an exposure or outcome that is unlikely to relate to the hypothesised causal mechanism, but which will include the same sources of bias or confounding as in the association of interest.[16] For example, lung cancer is a 'negative control' outcome not anticipated to be markedly influenced by levels of PA, but which is strongly related to confounding factors such as smoking. Therefore, an association observed between PA and lung cancer, which is similar to that observed between PA and CVD (where a causal mechanism has been hypothesised), would raise doubts about the validity of the latter effect estimate. Of note, non-CVD mortality was related to PA in the PURE study to a similar, if not larger, extent than CVD-related mortality, which potentially implies residual confounding in this context.

While demonstrating the utility of causal inference approaches, Kujala also highlights various limitations pertaining to residual confounding and other study biases, which cannot be fully accounted for by any individual method alone. Furthermore, he emphasises that when methodological flaws exist, increasing sample size does not always improve causal evidence (despite providing more precise estimates) and apparently robust associations may be subject to distortion through, for example, residual confounding.

While we would argue that Mendelian randomisation (MR)[17] is a powerful strategy for evading these problems, pleiotropy of the genetic instruments is a major consideration (although multiple strategies now exist to evaluate this), as is the current lack of genetic variants robustly associated with PA to serve as instrumental variables.[17] We, and others, have shown that BMI-associated genetic variants are also associated with measures of PA, suggesting a causal impact of adiposity on reducing levels of PA.[4 18] However, this further emphasises the confounding role of adiposity when evaluating associations between PA and longevity (as was emphasised by Kujala) and points to the need of identifying genetic variants that are directly associated with PA and not indirectly through other pathways. To date, the search for genetic variants associated with both self-reported and objective levels of PA has not been fruitful, although the availability of actigraphy data and other fitness metrics in large prospective studies such as the UK Biobank offer promise in this area.[19]

In practice, triangulation of a range of the outlined approaches, each with orthogonal key sources of bias,[20] alongside the improvement in the design of prospective studies and RCTs (eg, those within younger populations), individual participant meta-analyses,[21] co-twin control studies and MR are important steps in obtaining more reliable and generalisable estimates, confidence in findings and improving aetiological understanding in this area.

**OPEN ACCESS**

Check for updates

Linked

► http://dx.doi.org/10.1136/bjsports-2017-098639
► http://dx.doi.org/10.1136/bjsports-2018-099185

## REFERENCES

1  Kujala UM. Is physical activity a cause of longevity? It is notas straight forward as some would believe. Acritical analysis. *Br J Sports Med* 2018;52:914–8
2  Lear SA, Hu W, Rangarajan S, *et al*. The effect of physical activity on mortality and cardiovascular disease in 130 000 people from 17 high-income, middle-income, and low-income countries: the PURE study. *Lancet* 2017;390:2643–54.
3  Morris JN, Heady JA, Raffle PAB. Physique of London busmen. *The Lancet* 1956;268:569–70.
4  Richmond RC, Davey Smith G, Ness AR, *et al*. Assessing causality in the association between child adiposity and physical activity levels: a Mendelian randomization analysis. *PLoS Med* 2014;11:e1001618.
5  Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004;15:615–25.
6  Munafò MR, Tilling K, Taylor AE, *et al*. Collider scope: when selection bias can substantially influence observed associations. *Int J Epidemiol* 2017.
7  Äijö M, Kauppinen M, Kujala UM, *et al*. Physical activity, fitness, and all-cause mortality: an 18-year follow-up among old people. *J Sport Health Sci* 2016;5:437–42.
8  Newman AB, Dodson JA, Church TS, *et al*. Cardiovascular events in a physical activity intervention compared with a successful aging intervention: the life study randomized trial. *JAMA Cardiol* 2016;1:568–74.
9  Wing RR, Bolin P, Brancati FL, *et al*. Cardiovascular effects of intensive lifestyle intervention in type 2 diabetes. *N Engl J Med* 2013;369:145–54.
10  O'Connor CM, Whellan DJ, Lee KL, *et al*. Efficacy and safety of exercise training in patients with chronic heart failure: HF-ACTION randomized controlled trial. *JAMA* 2009;301:1439–50.
11  Anderson L, Oldridge N, Thompson DR, *et al*. Exercise-based cardiac rehabilitation for coronary heart disease: Cochrane systematic review and meta-analysis. *J Am Coll Cardiol* 2016;67:1–12.
12  Sagar VA, Davies EJ, Briscoe S, *et al*. Exercise-based rehabilitation for heart failure: systematic review and meta-analysis. *Open Heart* 2015;2:e000163.
13  Richard A, Martin B, Wanner M, *et al*. Effects of leisure-time and occupational physical activity on total mortality risk in NHANES III according to sex, ethnicity, central obesity, and age. *J Phys Act Health* 2015;12:184–92.
14  Davey Smith G. In: In Kuh D, Ben-Shlomo S, eds. *Socioeconomic differentials a life course approach to chronic disease epidemiology*. Oxford: Oxford University Press, 1997.
15  Millett C, Agrawal S, Sullivan R, *et al*. Associations between active travel to work and overweight, hypertension, and diabetes in India: a cross-sectional study. *PLoS Med* 2013;10:e1001459.
16  Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology* 2010;21:383–8.
17  Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014;23:R89–R98.
18  Schnurr TM, Viitasalo A, Eloranta AM, *et al*. Genetic predisposition to adiposity is associated with increased objectively assessed sedentary time in young children. *Int J Obes* 2017.
19  Klimentidis YC, *et al*. Genome-wide association study of habitual physical activity in over 277,000 UK Biobank participants identifies multiple variants including CADM2 and APOE. *bioRxiv* 2017.
20  Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *Int J Epidemiol* 2016;45:1866–86.
21  Ioannidis J. Next-generation systematic reviews: prospective meta-analysis, individual-level data, networks and umbrella reviews. *Br J Sports Med* 2017;51:1456–8.