

Onderzoek wearables niet bepaald vlekkeloos. Sloppy science



Tegenwoordig zie je steeds meer mensen sporten of anderszins actief bezig zijn terwijl ze een zogenaamde “wearable fitnessstracker” of kort gezegd een “wearable” dragen. Het zijn elektronische apparaten, die de polsfrequentie, het aantal afgelegde stappen, de verbruikte calorieën etc. bijhouden. Met enige regelmaat verschijnen er artikelen over deze apparaten waarin deze met elkaar vergeleken worden. De Stanford University in de Verenigde Staten rapporteert daar ook met enige regelmaat over. Niet altijd blijkt de academische opzet garant te staan voor een deugdelijke proefopzet. Zeer recent verscheen een artikel van de Stanford Universiteit waarin [de Apple Watch, Basis Peak, Fitbit Surge, Microsoft Band, Mio Alpha 2, PulseOn, en Samsung Gear S2](#) vergeleken werden ten aanzien van hartfrequentie en energieverbruik bij verschillende activiteiten. De conclusie van dat artikel was dat de hartfrequentie bij inspanning wel betrouwbaar gemeten kan worden met de fitness-sensoren, maar niet het calorieverbruik. Een kritische Amerikaan, Ray Maker, werkzaam in de IT-industrie, fanatiek sporter en eigenhandig tester van vrijwel alle “wearables” die op de markt gekomen zijn, laat op zijn eigen website zeer helder zien wat er allemaal schort aan de testen die in het Stanford-onderzoek gebruikt zijn. [Op de website DCRainmaker plaatste hij op 8 juni 2017 een artikel getiteld “Thoughts on the wearables studies \(including The](#)

[Stanford Wearables study\)](#)” . Hij laat zien dat de basis waarop de conclusies gebaseerd zijn, boterzacht is en dat conclusies op basis van de metingen met wearables zeer wankel zijn.

Niet volgens de handleiding

[In het beschreven Stanford-onderzoek werden frequent twee fitnesstrackers per pols van een proefpersoon gebruikt, dus in totaal vier.](#) Men deed dat om makkelijk met één proefpersoon in één sessie meetgegevens van vier apparaten te krijgen. Men was bang anders veel langer te moeten testen met veel meer proefpersonen. Op een foto bij het artikel van de Stanford University is dat ook te zien. Daarbij is het gelijk duidelijk dat meerdere apparaten niet op de plaats zitten die door de leverancier worden voorgeschreven om betrouwbare metingen te krijgen. Zo zit er één op het polsgewricht in plaats van 3 tot 8 centimeter hoger. Daarnaast werken hartfrequentiemeters die bij de pols gebruikt worden vaak op een elektro-optische basis. Zij kunnen op basis daarvan bij twee of meer per pols elkaar beïnvloeden(interferentie) en daardoor onbetrouwbare resultaten geven. Leveranciers weten dit zelf al en testen ze ook niet zo. Onderzoekers blijkbaar wel. Eén wearable per pols is het maximum en dan graag op de goede plek. Ook kunnen twee naast elkaar geplaatste apparaten aan één pols tegen elkaar stoten en zo mismetingen geven van het aantal stappen die door de proefpersoon gelopen heeft (met de motion-sensor.

Data

Ray Maker besteedt ook uitgebreid aandacht aan de wijze waarop data verkregen en bewerkt worden. Worden. Uit de beschrijving van de testen bleek bijvoorbeeld dat bij één van de apparaten, de Mio Alpha 2 de ruwe data niet rechtsreeks uit te lezen waren uit de sensor. Door een nogal aparte kunstgreep toe te passen werden de data toch in de studie opgenomen. Deze “wearable” stuurt wel data naar een app op een smartphone.

Door van de grafieken schermafdrucken te maken berekende men op basis daarvan één-minuut gemiddelden van de hartfrequentie. Bij het vergelijken van verschillende activiteiten(zitten, lopen, rennen en fietsen) lijken die activiteiten vloeiend achter elkaar aan ondernomen te zijn. In wezen is het één test, terwijl ze in het onderzoek als vier verschillende vermeld worden.

Calorie-verbruik

Het calorieverbruik kan uiteraard met “wearables” niet rechtstreeks gemeten worden. Daarvoor berekenen deze apparaten met algoritmen het vermoedelijke verbruik met als input de hartfrequentie die de “wearables” verschaffen. Elke leverancier zal daarbij zijn eigen systematiek gebruiken, waardoor het op voorhand geen betrouwbare uitslagen zal geven. Ray Maker laat ook zien dat de “wearables” geen schatting geven van het calorie-gebruik op basis van de hele activiteit waarover gemeten is, maar slechts op basis van een één-minuut-gemiddelde gedurende die activiteit.

Bij doorvragen over één apparaat dat de Stanford University gebruikte voor de calorie-verbruik-berekening, de PulseOn, bleek dat dit apparaat in tegenstelling tot de andere “wearables” niet de hartfrequentie als basis voor het caloriegebruik-algoritme hanteerde, maar de bewegings-sensor. De stappenteller dus. Zo vergelijk je appels met peren.

Sloppy science

Wat de Stanford University, maar ook andere onderzoeksinstellingen doen met “wearables” valt voor een groot deel te karakteriseren als “sloppy science”, slordige wetenschap. [Ik schreef over de Stanford University al eerder een artikel over het slordig wetenschap bedrijven met draagbare biosensors.](#) Ook in eigen land kan men er wat van. In het ReShape Center van het Radboud Universitair Medisch Centrum test men, onder leiding van Lucien Engelen, ook

“wearables” en schrijft men er artikelen over. Die moeten de apparaten blijkbaar een wetenschappelijke tintje geven. [De Chinese Check Me Health Monitor van het bedrijf Viatom die manchetloos bloeddruk zou moeten meten figureert in een wetenschappelijk artikel van ReShape-medewerkers dat bij kritische beschouwing slechts als “sloppy science” omschreven kan worden.](#)

Het trieste is dat de indruk ontstaat dat publiciteit en commercie belangrijker is dan de inhoud.

W.J. Jongejan