Please cite the Published Version

Additional Information: This is an Open Access article published in BMJ Health & Care Informatics.

Investigation of wearable health tracker version updates

Sandra I Woolley ![ORCID],[1] Tim Collins,[2] James Mitchell ![ORCID],[1] David Fredericks[1]

ABSTRACT

Background Wearable fitness trackers are increasingly used in healthcare applications; however, the frequent updating of these devices is at odds with traditional medical device practices.

Objective Our objective was to explore the nature and frequency of wearable tracker updates recorded in device changelogs, to reveal the chronology of updates and to estimate the intervals where algorithm updates could impact device validations.

Method Updates for devices meeting selection criteria (that included their use in clinical trials) were independently labelled by four researchers according to simple function and specificity schema.

Results Device manufacturers have diverse approaches to update reporting and changelog practice. Visual representations of device changelogs reveal the nature and chronology of device iterations. 13% of update items were unspecified and 32% possibly affected validations with as few as 5 days between updates that may affect validation.

Conclusion Manufacturers could aid researchers and health professionals by providing more informative device update changelogs.

INTRODUCTION

Wearable consumer activity monitors have substantially increased in popularity in the last decade,[1] and are increasingly used in healthcare applications and clinical trials. The US clinical trials database, ClinicalTrials. gov, returns 273 intervention results for 'Fitbit' (search accessed 30 May 2019) for a spectrum of studies investigating disease biomarkers, monitoring patient progress and incentivising lifestyle improvements. For example, studies include lifestyle interventions for overweight postpartum women (NCT03826394) and older adults at risk of cardiovascular disease (NCT03720327), a study with bowel disease patients to investigate biomarkers for predicting relapse (NCT03953794) and a study assessing rehabilitation progress of patients following knee surgery (NCT03368287).

The consumer fitness tracker market is fast-changing. New device models are regularly introduced and updated, and old models are retired. For example, the Garmin Vivosmart 'family' of wrist-worn activity trackers have included five physically distinct 'models': the (original) Vivosmart in 2014, Vivosmart HR in 2015 (which included optical heart rate sensing), Vivosmart HR+ in 2016, Vivosmart 3 in 2017 and Vivosmart 4 in 2018, all of which received several updates.

Wrist-worn trackers have increasingly supplemented step counting, activity monitoring, energy expenditure, sleep tracking and stress estimation with optical heart rate sensing from photoplethysmography sensors.[2] There is some debate about the reliability of these devices and their heart rate estimation,[3] and device validation studies have reached different conclusions for different health and exercise scenarios.[4–6] However, the devices can, and do, achieve significantly improved user activity behaviours and health outcomes.[7] These positive health effects have incentivised efforts towards new applications in corporate wellness,[8] health insurance[9] and in an increasing spectrum of clinical studies and patient-monitoring applications.[10 11] But, despite this move towards healthcare applications, device manufacturers are clear that their products are not medical devices. Indeed, the certification processes and validation timescales of medical devices are wholly at odds with the 'iterative characteristics' of consumer devices that can regularly and automatically update. At the launch of a pilot device manufacturer pre-certification programme aimed at addressing this gap, US Food and Drug Administration Commissioner Scott Gottlieb stated that '*Our method for regulating digital health products must recognize the unique and iterative characteristics of these products*'.[12] This 'iterative' nature also applies to device algorithms as manufacturers attempt to improve both parameter estimation and user satisfaction. So, not only does the appearance and behaviour of the devices update, but also the algorithms used in the logging and reporting of their data. These changes are made to the code that runs inside the processors embedded within these devices. In general, 'embedded

[1]School of Computing and Mathematics, Keele University, Keele, UK
[2]Department of Engineering, Manchester Metropolitan University, Manchester, UK

Correspondence to
Dr Sandra I Woolley;
s.i.woolley@keele.ac.uk

systems' are products or systems that contain embedded computer intelligence for purposes other than general-purpose computing. Today, most electronic products are embedded systems and, increasingly, in the age of the Internet of the Things (IoT), embedded system firmware updates can be communicated and applied automatically, without user intervention. Embedded code, being closer to the hardware of a system, is referred to as 'firmware'. Changes to device firmware can alter devices in fundamental ways. For example, by changing the rate and accuracy at which sensor signals are sampled, by changing the selection and filtering of signals, by changing algorithms that estimate measurements, such as heart rate and step count, and by selectively reporting and recording the results. From consumer goods, such as microwave ovens and washing machines, to mobile phones, cars and aeroplanes, devices can all be re-versioned with new firmware. Ideally, changes to firmware are recorded and itemised in a changelog document. Of course, manufacturers of commercial goods are under no obligation to share the details of their proprietary algorithms or reveal their intellectual property. Yet, at the same time, a level of open reporting can benefit users, stakeholders and, potentially, the manufacturers themselves.

It is noteworthy that, beyond firmware updates, there are additional software iteration complexities. IoT devices, such as wearables and smart home devices, can have their own operating systems and are often supported by cloud software services and interacted with via companion 'apps'. Updates to these other software components can also substantively impact device behaviour and data reporting.

The analysis of software code-related data and repositories is a mature field of research, but the focus has been on version control systems, such as GitHub,[13] open-source repositories and archives of user and developer fora.[14][15] There have been no analyses of consumer fitness tracking device repositories or changelogs. This may be due to several factors including the absence of source code and developer community engagements, the transient nature of device models or the relative sparsity of data in forum communications and device changelogs.

The neglect of updates in the literature has been reported by Vitale *et al*,[16] who observed, regarding the design of software updates, that '*no prior study can be found that investigated users' opinions regarding various design alternatives*'. In relation to operating system updates, Fagan *et al*[17] make several recommendations for improvement. For example, enabling updates to be reversible and decoupling security updates from other updates so that security updates can be made regularly, and other updates made selectively. They also recommend transparency to enable users to give consent to substantive changes. Beyond the academic literature, there are software developer and user experience (UX) designer opinions regarding best practice.[18–20] While style preferences vary, in general, the advice posited is (i) maintenance of a changelog, (ii) dating of updates, (iii) the grouping or labelling of

---

Box 1   Selection criteria for trackers

► Consumer-grade wrist-worn fitness tracker device designed for adult use.
► Includes heart rate sensing.
► Device family currently available.
► Model available for at least 12 months between 01 January 2017 and 01 April 2019.
► Manufacturer maintains a changelog.
► Device model specified in at least one clinical trial.

---

updates according to type or impact and (iv) making appropriate levels of details available to readers.

METHODS AND ANALYSIS

By searching the US Library of Medicine ClinicalTrials. gov database (search accessed 30 May 2019) with all device manufacturer names and models, we identified instances of named models of wearable heart rate sensing fitness trackers in patient studies. We then applied the device criteria listed in box 1 to select device models.

Changelogs were retrieved for tracker models meeting the selection criteria. Missing update release dates were extracted from manufacturer and user fora messages. Four researchers independently labelled each update item (updates typically comprise several items) according to type, specificity and potential to impact validation using the simple schema summarised in box 2. Differences in item labels were resolved by majority and arbitration. The type labelling scheme was based on the popular practice of added, changed, deprecated, removed, fixed and security labelling.[20] 'Bug fix' and 'issue resolution' items are frequently distinguished in update items, and so were provided with different labels. Similarly, 'addition' and 'improvement' changes were provided with distinct labels as were 'user interface change' items to

---

Box 2   Update type labels

► Bug fix
► Issue resolution
► Feature/function addition
► Feature/function improvement
► Algorithm adjustment
► User interface change
► Removal of items
► Security

Specificity Labels:
► Specified
► Semi-specified
► Unspecified

Estimated Potential to Impact on Validation Labels:
► Yes
► Possibly
► Probably not
► No

---

distinguish device presentation and interaction changes from other changes. A specific change label for 'algorithm adjustment' was used to reveal functional changes that more evidently affect data recording and validation, for example, '*Improvements to calculating resting heart rate*'.

RESULTS

Six device models met the selection criteria: the Fitbit Charge 2 (used in twenty-nine studies listed at ClinicalTrials.gov), Fitbit Charge 3 (used in four studies), Polar A370 (used in one study), Garmin Vivosmart 3 (used in two studies), Garmin Vivosmart 4 (used in one study) and Garmin Vivosport (used in one study). Only Polar's

changelog included dates. The update types, specificity and potential to impact validation for each of the six models are summarised in bar charts in figure 1. The chronology for types, specificity and potential to impact validation is illustrated in figure 2.

In total, the six device model changelogs comprised 194 update items. Overall, 13% of updates were unspecified and 32% possibly affected device validations. Maximum and minimum intervals between updates possibly affecting validations were 218 and 5 days, respectively.

Overall, user interface change updates accounted for 18% of all update items and the feature/function changes (additions and improvements) accounted for 33.5%.
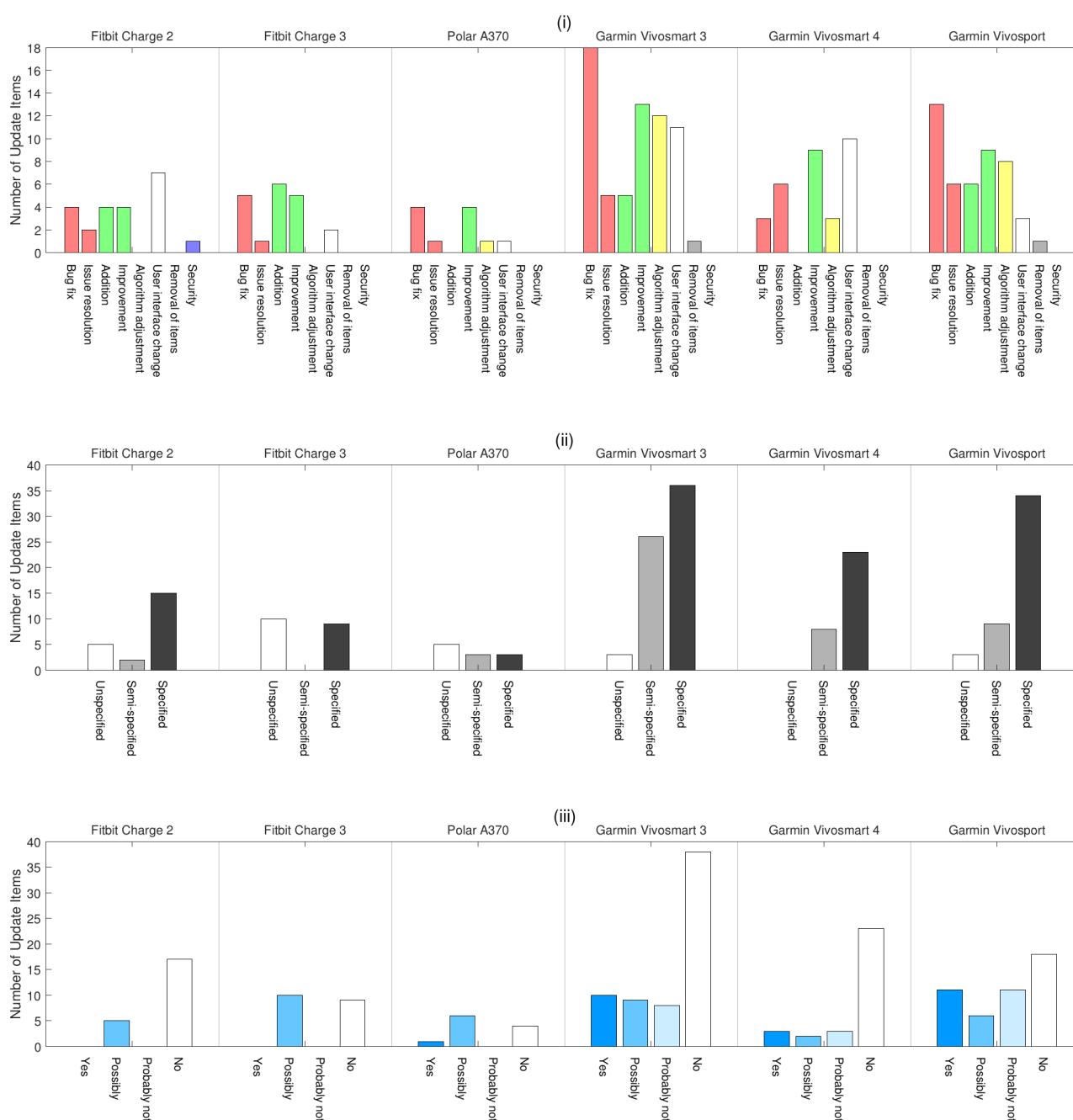

Figure 1 Update bar charts for each model for (i) type, (ii) specificity and (iii) estimated potential to affect validation.
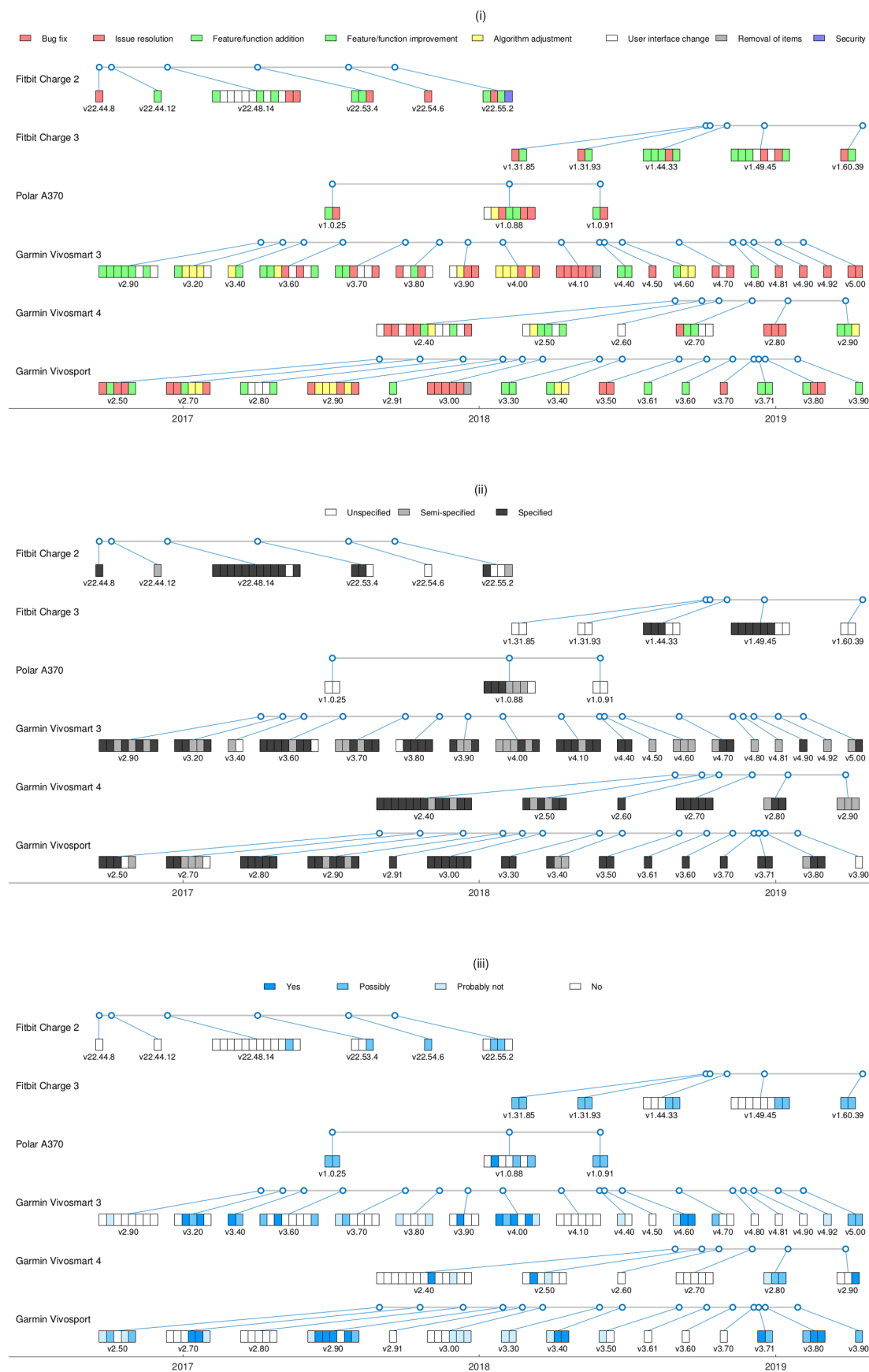
Figure 2 Update chronology for (i) type, (ii) specificity and (iii) estimated potential to affect validation.

Together, bug fixes and issue resolutions accounted for 35% and the percentage of algorithm adjustment updates was 12%. Only one update was identified as a security update and only two updates were identified as item removals.

DISCUSSION

Although ClinicalTrials.gov returned 273 intervention results for 'Fitbit', only 38 studies satisfied the selection criteria. Many studies were excluded because they did not meet the adult heart rate sensing or device availability criteria, or because they failed to identify device models. It is noteworthy that activity tracking studies cannot easily be enumerated because there is no labelling schema and there is a lack of consistency in terminology. For example, studies refer variously and ambiguously to devices by manufacturer name alone, for example, 'Fitbit', or generically as 'activity tracker', 'activity monitor', 'smartwatch', 'wearable sensor', and so on. We identified trials meeting our selection criteria by searching for all manufacturer and model names. But, ideally, trials would be labelled according to some useful schema.

Of course, researchers may do well to avoid specific device details at the proposal stage of their clinical trials because, by the time of recruitment, the specified models may be unavailable and newer and more functional models may be available. But, at the same time, trial documentation would ideally provide a more meaningful level of information about study materials. Perhaps, if trial devices can be upgraded and updated post-proposal, then trial documentation should be updated also? Perhaps the consequence of 'iterative' devices is the need for iterative documentation?

All the device changelogs in our study contained some entries that were difficult or impossible to confidently classify. Of course, device updates may fit several categories, for example, a bug fix could be a security update and a new feature could be an interface change, and so on. But it should still be possible to label and specify items. Common unspecified items included 'various other updates' and 'bug fixes' which do not indicate the number of changes and their potential to impact recorded data. Ambiguous entries like these may not be considered problematic for consumer-grade devices but, when the same devices are used for scientific research, the impact of an update on device performance and validation could be significant. As manufacturers increasingly promote the use of devices in research and health-related applications, there may be more incentives for the maintenance of accurate and unambiguous changelogs.

Alongside the need for accurate changelogs and study descriptions, there is a wider need to more accurately report device models *and* version numbers in the literature.[21] It should be important to verify that devices are used at the same version that they were validated at, and that the version number is reported in the interests of repeatability. Version numbers for statistical analysis software are often reported for this reason, but the reporting of device firmware versions is almost always neglected. Because devices can update automatically during data synchronisation, it is likely that firmware updates are often applied inadvertently in the middle of studies without the researcher noticing or reporting the change.

Ideally, there would be a better awareness of device versioning, and device versions would be controlled or at least reported in study and research literature. Manufacturers could also benefit the community by providing accessible changelog details and labelling appropriate for research and healthcare users and, perhaps, supported by simple visual bar chart and chronology summaries similar to those reported in figures here.

CONCLUSIONS

We observed striking differences between manufacturer changelog practices. For example, differences in level of detail, in voice and presentation style, and in the inclusion and omission of date information. Ideally, manufacturers would adhere to informative and consistent (and, perhaps, standardised) changelog formats that provide useful and accessible information to clinical researchers and healthcare professionals. For example, by labelling updates according to type, highlighting key functional/algorithm change updates and avoiding unspecified entries and ambiguous update descriptions. There is also a need for version reporting in the wearable monitoring literature and an improved awareness that firmware updates can nullify device validations.

ORCID iDs
Sandra I Woolley http://orcid.org/0000-0002-7623-2866
James Mitchell http://orcid.org/0000-0001-6051-2567

REFERENCES
1 Mintel Press Office. Sales of fitness bands and smartwatches approach four million devices. online, 2018. Available: http://www.mintel.com/ press-centre/technology-press-centre/sales-of-fitness-bands-and-smartwatches-approach-four-million-devices [Accessed Mar 2018].
2 Henriksen A, Haugen Mikalsen M, Woldaregay AZ, *et al.* Using fitness Trackers and Smartwatches to measure physical activity in research:

analysis of consumer Wrist-Worn Wearables. *J Med Internet Res* 2018;20:3.

3 Oniani S, Woolley SI, Pires IM, *et al*. Reliability assessment of new and updated consumer-grade activity and heart rate monitors. *IARIA SensorDevices* 2018.

4 Baron KG, Duffecy J, Berendsen MA, *et al*. Feeling validated yet? A scoping review of the use of consumer-targeted wearable and mobile technology to measure and improve sleep. *Sleep Med Rev* 2018;40:151–9.

5 Stahl SE, An H-S, Dinkel DM, *et al*. How accurate are the wrist-based heart rate monitors during walking and running activities? are they accurate enough? *BMJ Open Sport Exerc Med* 2016;2:e000106.

6 O'Driscoll R, Turicchi J, Beaulieu K, *et al*. How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies. *Br J Sports Med* 2018:bjsports-2018-099643.

7 Whelan ME. *Persuasive digital health technologies for lifestyle behaviour change*. Doctoral dissertation, Loughborough University, 2018.

8 Giddens L, Leidner D, Gonzalez E. The role of Fitbits in corporate wellness programs: does step count matter? *In Mobile Apps and Wearables for Health Management, Analytics, and Decision Making. AIS*, 2017:3627–35.

9 Paluch S, Tuzovic S. Leveraging pushed self-tracking in the health insurance industry: how do individuals perceive smart wearables offered by insurance organization? *In European Conference on Information Systems. AIS*, 2017:2732.

10 Chiauzzi E, Rodarte C, DasMahapatra P. Patient-Centered activity monitoring in the self-management of chronic health conditions. *BMC Med* 2015;13:77.

11 Mercer K, Giangregorio L, Schneider E, *et al*. Acceptance of commercially available wearable activity trackers among adults aged over 50 and with chronic illness: a mixed-methods evaluation. *JMIR mHealth and uHealth* 2016;4.

12 FDA. Fda selects participants for new digital health software precertification pilot program. online, 2017. Available: https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm577480.htm [Accessed May 2019].

13 Kalliamvakou E, Gousios G, Blincoe K. The promises and perils of mining GitHub. *In Proceedings of the 11th working conference on mining software repositories. ACM*, 2014:92–101.

14 Chen T-H, Thomas SW, Hassan AE. A survey on the use of topic models when mining software repositories. *Empir Software Eng* 2016;21:1843–919.

15 Khalid H, Shihab E, Nagappan M, *et al*. What do mobile APP users complain about? *IEEE Software* 2015;32:70–7.

16 Vitale F, Mcgrenere J, Tabard A. High costs and small benefits: a field study of how users experience operating system Upgrades. *In Proceedings of the CHI Conference on Human Factors in Computing Systems. ACM*, 2017:4242–53.

17 Fagan M, Khan MMH, Nguyen N. How does this message make you feel? A study of user perspectives on software update/warning message design. *Hum. Cent. Comput. Inf. Sci.* 2015;5.

18 Gill R. As a designer I want better release notes, 2017. Available: https://uxdesign.cc/design-better-release-notes-3e8c8c785231 [Accessed May 2019].

19 Chen J. Outstanding release notes examples (and how to use each), 2017. Available: https://www.appcues.com/blog/release-notes-examples [Accessed May 2019].

20 Lacan O. Keep a Changelog. Available: https://keepachangelog.com/en/1.0.0/ [Accessed May 2019].

21 Collins T, Woolley SI, Oniani S, *et al*. Version reporting and assessment approaches for new and updated activity and heart rate monitors. *Sensors* 2019;19.