Daily & hourly adherence: towards understanding activity tracker accuracy.

Abstract
We tackle the important problem of understanding the accuracy of activity tracker data. To do this, we introduce the notions of daily and hourly adherence, key aspects of how consistently people wear trackers. We hypothesise that these measures provide a valuable means to address accuracy problems in population level activity tracking data. To test this, we conducted a semester-long study of 237 University students: 88 Information Technology, 149 Medical Science. We illustrate how our adherence measures provide new ways to interpret data and valuable insights that take account of tracker data accuracy. Finally, we discuss broader roles for daily and hourly adherence measures in activity tracker data.

Author Keywords
Activity Tracker; Data Accuracy; Adherence; Fitbit

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous

Introduction
Using activity trackers to improve health is promising [2]. However, to understand patterns of use of trackers, we need to understand the accuracy of the data. One key contribution to inaccuracy follows from the fact that even the
The most motivated user does not wear their tracker all the time over years, months or even weeks. To really understand how active people are, one must account for the fact that inconsistent wear gives an incomplete picture. To address this, we introduce the notion of adherence to capture key aspects of the level of wear and show how to use this to give a more accurate picture.

Daily adherence is the percentage of users who wore their trackers each day. Figure 1 illustrates this for a population of 237 people over a 2 months period. The figure shows an overall steady drop and also cyclic patterns. Previous literature has reported overall drop-out rates, the rate at which people stop wearing their trackers [9, 16]. However, that body of work ignores the accuracy of tracking data during periods when people did wear their tracker.

Hourly adherence is the number of hours users wore their trackers on days they remembered to put them on. It is important to consider this in addition to daily adherence because it reveals how valid the tracking data is on each day. We took the term adherence from its use in medical intervention research [18] where the data from study participants is only used if they wore the device the required number of hours on each of the required number of days. Rather than simply use adherence as an exclusion criterion, we now show how to use it to make sense of the data that is available from the many people who elect to use trackers as part of their normal lives. This data has the potential to give valuable understanding of populations of users.

While there is a growing body of work on non-medical, general use of activity trackers [16, 12, 6, 11] we have found no reports of work that considers daily and hourly adherence to tackle the challenge of data inaccuracy.

We hypothesize that:

- Daily and hourly adherence is important for the accurate interpretation of long term physical activity tracker data.

To explore the power of these notions, we conducted a semester long study with 237 university students: IT (88) and Medical Science (149). Each was given a Fitbit Zip device and we analysed the data to determine their daily and hourly adherence.

We are the first to report on the daily and hourly adherence levels and patterns on a large group of students. We show how analysis based on daily adherence discloses interesting similarities and differences between the two student groups. This highlights the way that daily adherence has the potential to be a significant source of inaccuracy that can differ across populations. We also show how hourly adherence analysis complements and adds to the picture and needs to be considered as a source of inaccuracy for simplistic analyses of tracker data.

Background
This section first explores the nature of accuracy in the context of physical activity tracking. We then explain the previous use of adherence in such data. Next we show how these aspects link to the main studies of activity tracker use. The section concludes with the positioning of our study to address gaps in the literature.

A number of barriers have been reported in the study of activity tracker adoption including motivation [16, 11, 7], aesthetics [16, 12, 6], maintenance challenges [14] and accuracy. Our work concerns the last of these. Reflecting its importance, there has been work to understand its forms and impact. We summarise this in Table 1, distinguishing three categories of accuracy challenges. Our work...
tackles the third, presentation and comprehension, which refers to problems due to data being misrepresented or misunderstood. Consolvo used the calorie count presented in many health applications as an example where users are not aware that this value is really an approximation [7, p. 230]. Moreover, many applications present graphs and summaries which either ignore or do not indicate that there is missing data [7, p. 234]. This is a problem for personal tracker data as it is likely to be incomplete over the long term [16, 7]. Yang et al. [19] reported that users often incorrectly interpreted the inaccuracy in tracking data. User’s daily and hourly adherence data is therefore very important to understand as it directly impacts the accuracy and ultimately presentation and comprehension of the data.

Adherence is used in medical literature on pedometer intervention and health research [3, 8, 4]. For example, such work reports using only data where participants wore the device for at least 10 hours a day; any less than this was considered too inaccurate to use. This is a good approach to accuracy in measuring intervention outcomes. This is quite different from work on normal use of trackers.

Medical literature includes studies of special user populations. For example, Cadmus-Bertram et al. [4] reported on a 16-week Fitbit pedometer intervention study with 25 overweight or obese, postmenopausal women. They showed that the median participant wore her tracker for at least 10 hours on 95% of days with no significant decline over time. We have found no reports of such hourly adherence for a broader user population.

Studies of activity tracker use have reported what they call a drop-off patterns which describes how long a user wore their tracker before it is abandoned. For example, Shih et al. studied 26 undergraduate students over 6 weeks and found that 65% of participants had dropped off after just 2 weeks [16]. Moreover, based on survey data, Endeavour partners reported that more than a third of the owners of smart wearables had abandoned them after 6 months [9]. However, this work on drop-off rates ignores the level of adherence during use and the many ways that people may want to use trackers [6, 12, 14].

There are good reasons to expect daily and hourly adherence to differ across time. For example, a 7-day study of university students [17] reported lower levels of physical activity on weekends. Daily adherence analysis is needed to account for this.

In summary, there is a growing body of work to gain understanding of the ways that people make use of physical activity trackers. In particular, there has been study of drop-off and of the factors affecting both use and drop-off. But, outside medical intervention studies, adherence has not been considered. Yet it seems to have an important role for understanding the ways populations of users actually make use of the devices, giving a more nuanced view than pure drop-off but also pointing to patterns across the week and over long periods of time.

### Study Design

To explore our hypothesis, that daily and hourly adherence is important for understanding the accuracy of long term physical activity tracker data, we designed a study that collected long term data for two populations of users. Studying daily and hourly adherence across these populations gave us the opportunity to see whether these measures disclosed interesting similarities and differences that impact accuracy. We now describe the populations and the procedures.

We recruited 237 students from 2 university courses: 88 information technology (IT) and 149 medical science (MED).
We expected that these students would have different attitudes to activity and tracking. The MED students were second year undergraduates. Their formal studies encourage them to be conscious of health benefits of physical activity. The IT students were third year undergraduates in an HCI subject whose classrooms are at a different part of the university campus. Their studies do not have a health focus. However, the HCI subject had a theme on physical activity, treated in a lecture and homework reading [5, 13] and their main assignment was to design a user interface to promote physical activity and reduce inactivity. These groups allowed us to observe adherence differences from students in different social and physical environments.

The Fitbit Zip was provided, on loan for the semester, to each student for the duration of the study. We chose these because they were low cost and had up to 6 months of battery life avoiding maintenance barriers reported in other work [14]. Per-minute steps data was obtained through the Fitbit Rest API which allowed us to determine detailed adherence patterns.

### Results

In this section, we report our analysis of the adherence patterns and how these give insights about accuracy of the data sets. We first present and discuss population level daily adherence patterns and discuss our analysis of distinctive features. We then present hourly adherence levels and show how they extend the picture emerging from daily adherence. Finally, we present drop-out patterns, using our data to replicate [16] and highlighting how our adherence measures give important new insights into accuracy.

#### Daily Adherence

Figure 2 presents daily adherence of IT (the lower blue line) versus MED students (the upper red line) over the study period. The figure shows weekly cycles through the teaching weeks of the semester. We have labelled the semester break; this is flatter than other weeks. Table 2 compares means of both adherence measures for weekdays and weekends over the full study. Table 3 does this for weekdays in the teaching weeks, compared with the mid-semester break.

Figure 2 is a striking demonstration of the way that our daily adherence measure highlights weekly patterns of peaks and troughs. This is remarkably consistent across both student groups. This pattern led us to compare the overall mean daily adherence on weekdays and weekends. This is summarised in the upper part of Table 2. For both student groups, there was lower daily adherence on weekends. For IT students this was 19% on weekends versus 28% (p=0.02, 2 sample t-test) on weekdays. A similar pattern, albeit at a high level of daily adherence applied for the MED students, who had 39% on weekends and 48% on weekdays.

### Table 2: Mean daily and hourly adherence: weekdays versus weekends.

<table>
<thead>
<tr>
<th></th>
<th>IT</th>
<th>MED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekdays</td>
<td>28%</td>
<td>48%</td>
</tr>
<tr>
<td>Weekends</td>
<td>19%</td>
<td>39%</td>
</tr>
<tr>
<td>p-value</td>
<td>0.02</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

### Table 3: Mean daily and hourly adherence: weekdays in the teaching weeks versus the mid-semester break.

<table>
<thead>
<tr>
<th></th>
<th>IT hr</th>
<th>MED hr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekdays</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>Break</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;0.001</td>
<td>0.01</td>
</tr>
</tbody>
</table>

The Fitbit Zip was provided, on loan for the semester, to each student for the duration of the study. We chose these because they were low cost and had up to 6 months of battery life avoiding maintenance barriers reported in other work [14]. Per-minute steps data was obtained through the Fitbit Rest API which allowed us to determine detailed adherence patterns.

#### Hourly Adherence

We expected that these students would have different attitudes to activity and tracking. The MED students were second year undergraduates. Their formal studies encourage them to be conscious of health benefits of physical activity. The IT students were third year undergraduates in an HCI subject whose classrooms are at a different part of the university campus. Their studies do not have a health focus. However, the HCI subject had a theme on physical activity, treated in a lecture and homework reading [5, 13] and their main assignment was to design a user interface to promote physical activity and reduce inactivity. These groups allowed us to observe adherence differences from students in different social and physical environments.

The Fitbit Zip was provided, on loan for the semester, to each student for the duration of the study. We chose these because they were low cost and had up to 6 months of battery life avoiding maintenance barriers reported in other work [14]. Per-minute steps data was obtained through the Fitbit Rest API which allowed us to determine detailed adherence patterns.

#### Results

In this section, we report our analysis of the adherence patterns and how these give insights about accuracy of the data sets. We first present and discuss population level daily adherence patterns and discuss our analysis of distinctive features. We then present hourly adherence levels and show how they extend the picture emerging from daily adherence. Finally, we present drop-out patterns, using our data to replicate [16] and highlighting how our adherence measures give important new insights into accuracy.

#### Daily Adherence

Figure 2 presents daily adherence of IT (the lower blue line) versus MED students (the upper red line) over the study period. The figure shows weekly cycles through the teaching weeks of the semester. We have labelled the semester break; this is flatter than other weeks. Table 2 compares means of both adherence measures for weekdays and weekends over the full study. Table 3 does this for weekdays in the teaching weeks, compared with the mid-semester break.

Figure 2 is a striking demonstration of the way that our daily adherence measure highlights weekly patterns of peaks and troughs. This is remarkably consistent across both student groups. This pattern led us to compare the overall mean daily adherence on weekdays and weekends. This is summarised in the upper part of Table 2. For both student groups, there was lower daily adherence on weekends. For IT students this was 19% on weekends versus 28% (p=0.02, 2 sample t-test) on weekdays. A similar pattern, albeit at a high level of daily adherence applied for the MED students, who had 39% on weekends and 48% on
weekdays (p=0.03, 2 sample t-test).

Similarly, Figure 2’s flat section at the mid-semester break (30-Sep to 4-Oct) motivated scrutiny of that week, compared to weekdays in teaching weeks. This shows that daily adherence levels on weekdays of the mid-semester are rather like weekends in the teaching semester. The upper part of Table 3 shows that the weekdays in the teaching semester have far higher daily adherence than the break weekdays. The daily adherence in the semester break did not look like the weekends. The means for all weekends and the mid-semester break weekdays (IT 19% vs. 16% p>0.05, MED 36% vs. 39% p>0.05 , 2 sample t-test). and hourly adherence (IT 11.4 vs. 10.3 hours p>0.05, MED 11.8 vs. 11.6 hours p>0.05).

These results support our hypothesis for daily adherence and indicate that accuracy of activity data must take account of the times, such as weekends and also our semester breaks. These results on daily adherence extend the previous findings on lower activity levels over weekends [1] to involve lower daily adherence levels as well.

**Hourly Adherence**

Figure 3 shows the IT (lower blue) and MED (upper red) hourly adherence rates across the study period. This data allowed us to make 2 observations relating to accuracy.

First, many students failed to reach the threshold (10 hours of wear or more) considered valid for medical intervention studies [1, 4]. Table 4 reports the percentage of days where the median participant reached 10 hours or more. MED students reached this level on 75% of days and for IT students it is only 35%. Also, while IT students were less adherent than MED students, they did wear trackers for at least 6 hours on 85% of days shown in the table. The population level hourly adherence data in Figure 3 also shows this quite consistent adherence above 6 hours. We note that the standard deviation levels are high, between 4 and 5 hours, reflecting the large variation within student groups.

Second, IT students wore their trackers for fewer hours than MED students, with means of 9.5 hours and 11.5 hours per day respectively (p<0.001, 2 sample t-test). Lower data in Table 2 shows that MED students wore their trackers for longer on weekdays than weekends. This was not the case for IT students who had similar means of 9.4 hours on weekdays and 9.7 hours on weekends. Combining hourly and daily adherence, we see that while daily adherence is lower on weekends, this was not so for IT students for hourly adherence.

These results support our hypothesis for the importance of hourly adherence in understanding the accuracy of activity tracker data. Moreover, there is no need to discard data.
Figure 4: Participant drop-out rate of IT (blue) and MED (red) students over the study period. N=237

as is the practice in medical research. Instead adherence measures can still make population data useful. This makes it feasible to draw on the large amounts of data from many users whose data can still provide valuable insights on real uses of activity trackers.

Drop-off Rate
Drop-off rate is a cumulative measure of the percent of users who stopped using their tracker. We replicated the drop-off analysis in [16], with the results shown in Figure 4. Our student population had lower drop-off than in [16] with 35% drop-off compared to 75% after 30 days. In addition, it took 45 days for IT students to reach 75% and even longer for MED students. This may be due to a number of factors including intervention effects, the populations and environment. It points to the need for further studies for different populations. Studying drop-off rates only indicates when students stopped wearing their trackers. However, it completely omits the effects we have described above.

Limitations
Our study covered 51 days. Different effects may emerge in the longer term. While we had a large population compared to many studies, it is distinctive and can be best seen as adding to the understanding of tracker data. Also, our Fitbit Zip device has limitations, such as limited waterproofing and how it can be worn (i.e., clip on). Notably, the Fitbits were on loan only for the study and this is likely to have impacted results compared with other populations such as those who own their devices.

Conclusion & Future Work
We conclude that daily and hourly adherence measures are important. Daily adherence varied significantly at different times such as weekends and the mid-semester break and this impacts accuracy of the population data. Hourly adherence proved to be a potential source of inaccuracy as many failed to wear their tracker for extended hours. Our results point to a need for further work on the perceived inaccuracy at the individual level. Combining population and individual level insights has the potential to offer an individual a clearer picture of their own data accuracy. They can help us better determine how and when to apply interventions and also tailoring applications to individual patterns such as sending reminders on weekends to increase adherence. Accounting for adherence also has the potential to inform design of better interfaces for long term activity tracker data. Notably, visualisations can show exact daily and hourly adherence in longer term data to help them appreciate what their data represent.

Acknowledgements
Special thanks to Dr Tina Hinton facilitating with students. This research was supported by funding from the Faculty of Engineering & Information Technologies, The University of Sydney, under the Faculty Research Cluster Program.
References


[4] Lisa Cadmus-Bertram, Bess H Marcus, Ruth E Patterson, Barbara A Parker, and Brittany L Morey. 2015. Use of the Fitbit to Measure Adherence to a Physical Activity Intervention Among Overweight or Obese, Postmenopausal Women: Self-Monitoring Trajectory During 16 Weeks. *JMIR mHealth and uHealth* 3, 4 (2015), e96. DOI: http://dx.doi.org/10.2196/mhealth.4229

[5] Timothy S Church and Steven N Blair. 2008. When will we treat physical activity as a legitimate medical therapy...even though it does not come in a pill? *British journal of sports medicine* 43, 2 (2009), 80–81. DOI: http://dx.doi.org/10.1136/bjsm.2008.053850


