Appendix A. CGM Dataset

We include information on the range of values for each external factor in our dataset in Table 4. Note that all extreme weather events and all temporal events are binary-valued.

We plot the distribution of number of days recorded for each individual in Figure 4. The large spike in the last bucket contains roughly 10% of individuals in our cohort, and represents people who have recorded CGM data on over 75% of days over the 2.5 year duration of data collection.

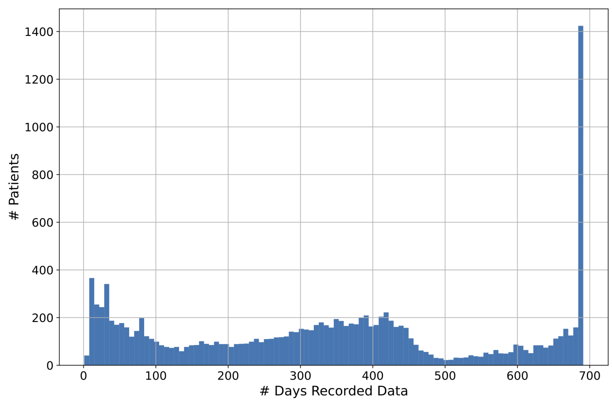

Figure 4: Distribution of number of recorded days of CGM data per individual. The spike in the last bucket is due to about 10% of individuals providing recordings (nearly) every day over the 2.5-year data collection period.

Appendix B. Classifier Details

Information on the values input into the random forest classifiers is provided in Table 5. Note that the CGM and activity data is not used for classifiers that use no CGM or activity data features.

Appendix C. Additional Results

Here we provide plots showing the PR-AUC performance of our classifiers. We observe similar results as described in Section 3.2.

In Figure 5, we plot the PR-AUC of models across 8 of the 12 external factors studied (the remaining 4 were excluded due to insufficient data for evaluation). This plot corresponds to Figure 2 where we showed ROC-AUC performance. We note similar results – we are able to consistently outperform a random baseline (here the random baseline is the prevalence).

In Figure 6, we plot PR-AUC of models predicting the effect of TIR while modulating the number of days of CGM and activity data used to compute the CGM and activity summary features. This plot corresponds to Figure 3 where we evaluated the models with respect to ROC-AUC. We observe similar results here, with models generally improving PR-AUC score when more days of data are used to generate the summary statistics for CGM and activity data.

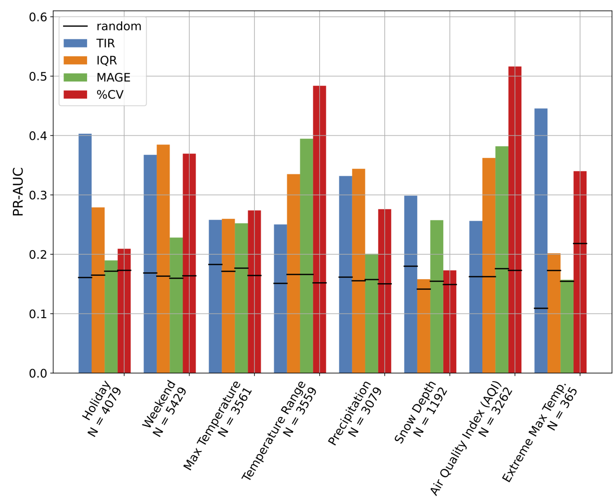

Figure 5: Performance (PR-AUC) of models at predicting the effect of external factors on individuals. Each bar represents a classifier. External factors that have < 1000 datapoints (days of data) for training and evaluation combined are excluded.

Table 4: Summary of external events and the range of values that occur in our filtered cohort. 1^{st} and 99^{th} percentile values are computed from our filtered cohort (1^{st} and 99^{th} percentile are used as opposed to min and max as all continuous factors have significant outliers). Extreme weather and temporal events are all binary valued. *AQI is on a scale from 0 to 500.

Category	External Factor	1^{st} Percentile	99^{th} Percentile
Weather	Max Temperature (˚C)	-6.7	38.3
	Temperature Range (˚C)	0	23.3
	Precipitation (mm)	0	46
	Snow Depth (mm)	0	229
	AQI*	10	130
Extreme Weather	Max Temperature	0	1
	Temperature Range	0	1
	Precipitation	0	1
	Snow Depth	0	1
	AQI	0	1
Temporal Events	Weekends	0	1
	Holidays	0	1

Table 5: Features used for Random Forest classifiers.

Feature Category	Summary of features used
Demographics	We use features like age and gender here. We also compute summary statistics (mean, standard deviation, and 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentiles) for each environment condition (weather and AQI) in the region the individual lives in.
Medical	We use over 100 standard metrics like most recent BMI and HbA1C.
CGM and Activity Data	Daily summary statistics are computed including mean, median, max, min, mean of derivative, and the standard deviation of derivative. These statistics are then summarized per individual by computing the median across all days used for the feature (i.e., first 3 days, first 10 days, or all days available). These features are not used when no CGM or activity data are allowed.

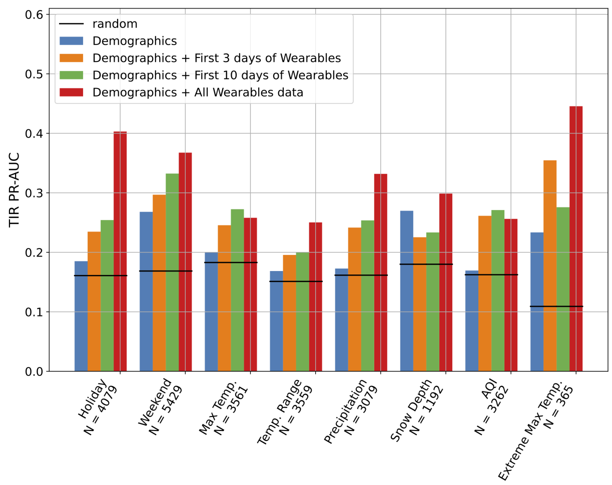

Figure 6: Performance increases given additional CGM and activity data. Showing PR-AUC of models predicting TIR effect on individuals. External factors that have < 1000 datapoints (days of data) for training and evaluation combined are excluded.