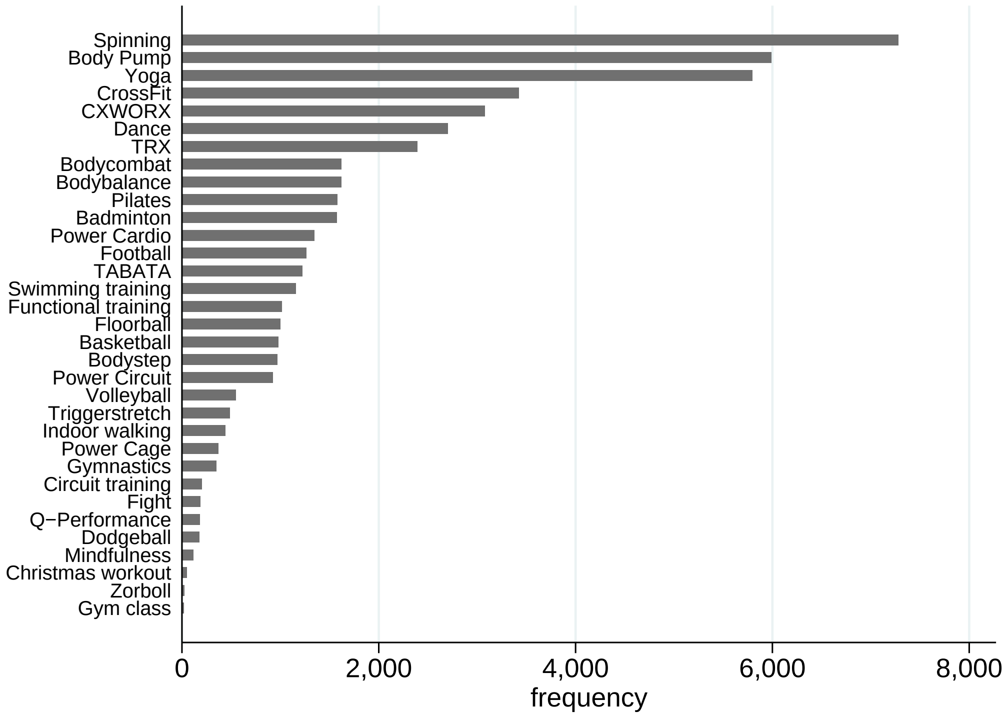
Online Appendix: Additional material

Figure 1: Visits per class



Classes are categorized into the 6 main categories as follows: 1 (endurance and strength): Bodycombat, Bodystep, Cirkelgym, CrossFit, Fight, Fys, Gymclass, Gympa, Krafttag, Powercage, Powercircuit, Tabata; 2 (body and mind): Bodybalance, Mindfulness, Pilates, Triggerstretch, Yoga; 3 (cycle/run/swim): Spinning, Indoor walking, Swim-training; 4 (core): Cxwork, Functional training, Q-performance, TRX; 5 (ball sports): Badminton, Basketball, Football, Indoor hockey, Dodgeball, Volleyball; 6 (dance): Dancing classes.

Figure 2: Distribution of dates of buying contract for control and treatment groups

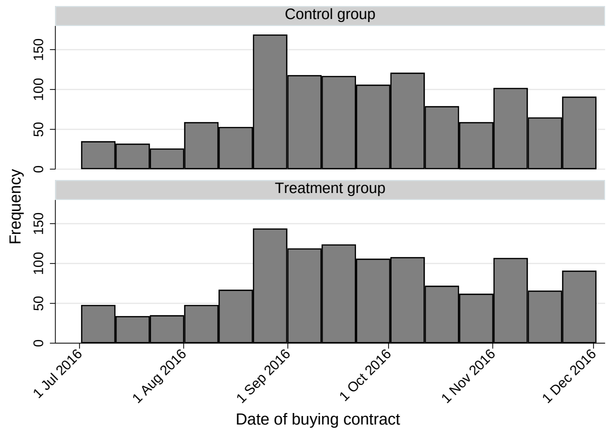
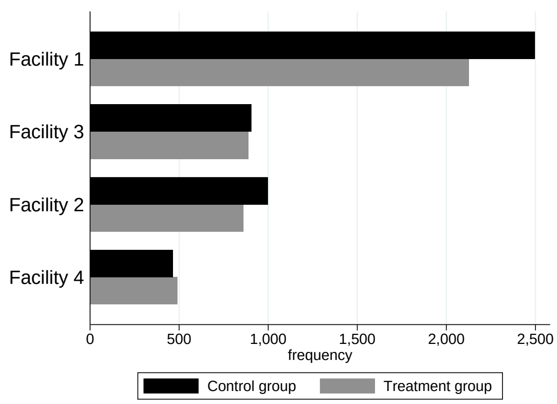
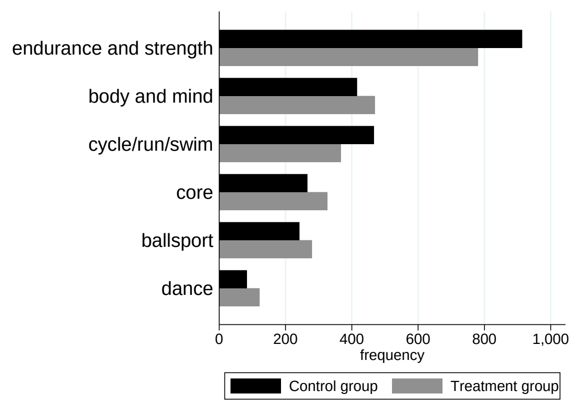


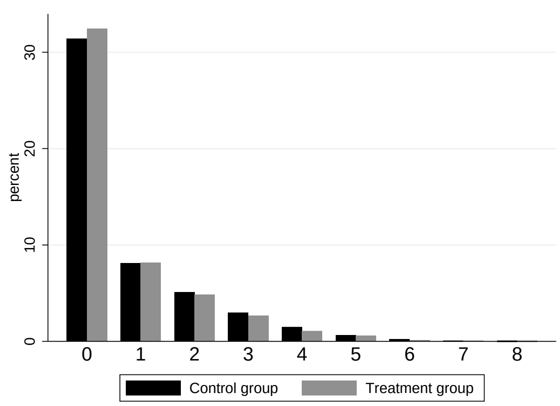
Figure 3: Comparing control and treatment: pre-experiment gym attendance



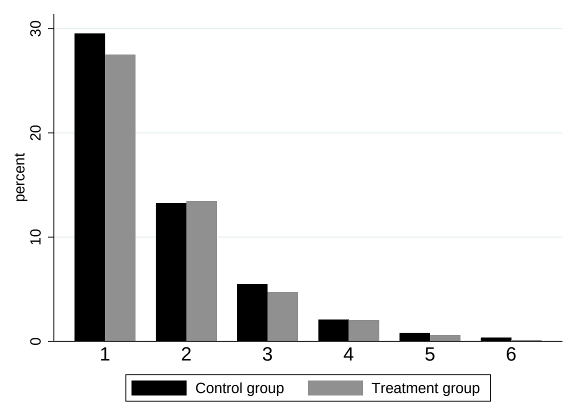
(a) Free training per facility



(b) Group classes



(c) Distribution of weekly free training visits



(d) Distribution of weekly class training visits, excluding zeros (88 % of all observations)

Figure 4: Example of an email reminder

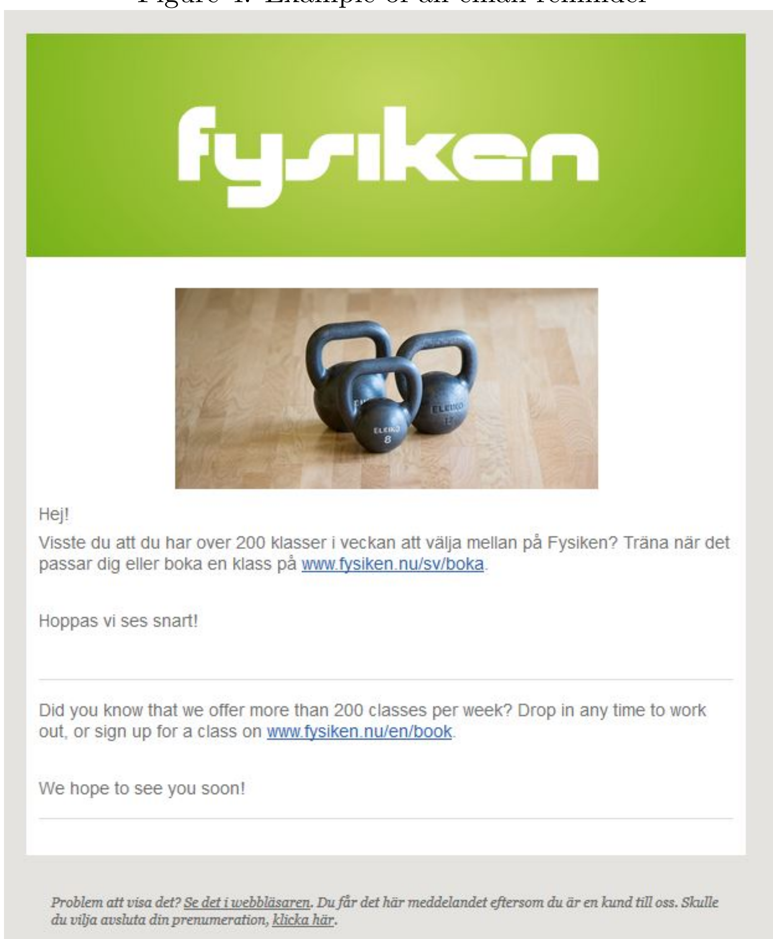
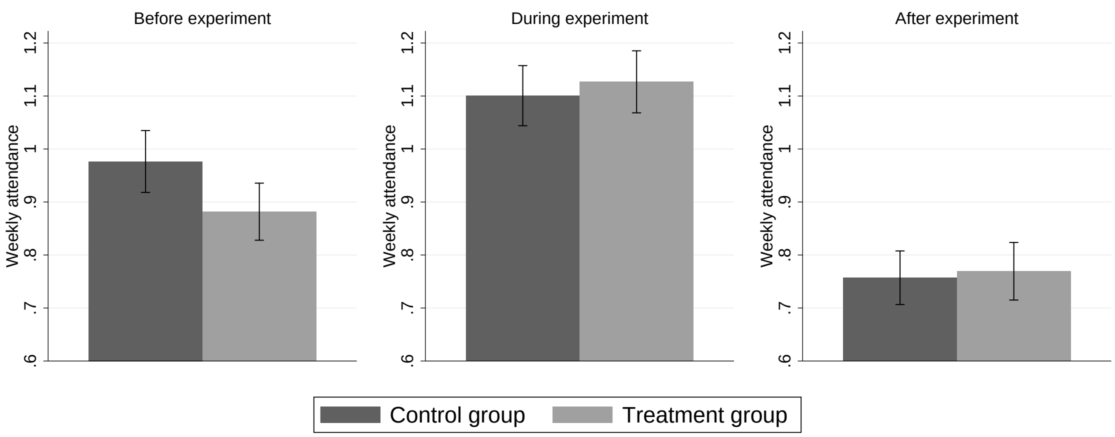
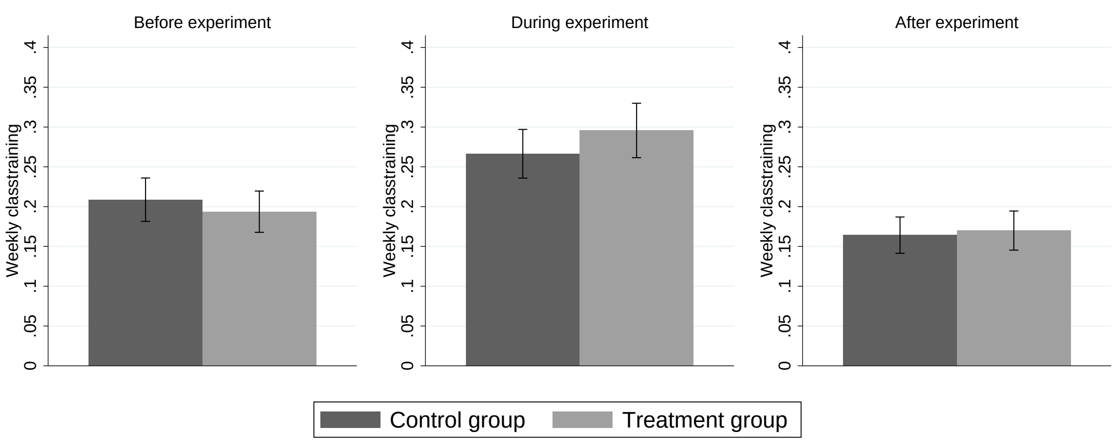


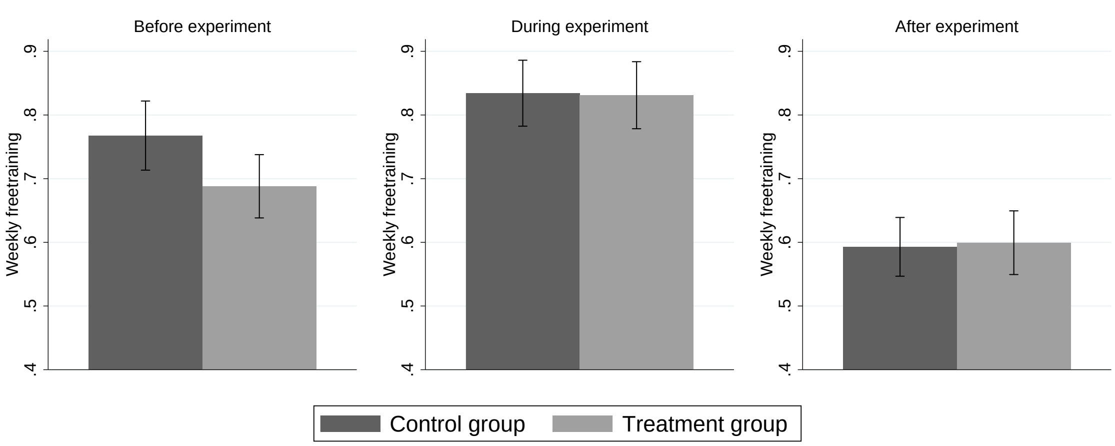
Figure 5: Weekly gym attendance: group means (with 95 % confidence interval)



(a) Total visits (DiD estimate during: 0.12, DiD estimate post: 0.11)



(b) Class training (DiD estimate during: 0.05 : DiD estimate post: 0.02)



(c) Free training (DiD estimate during: 0.08, DiD estimate post: 0.09)

Table 1: Regressions on weekly gym attendance (linear models)

	DiD model		
	(1)	(2)	(3)
	Visits	Class training	Free training
Experiment (during)	0.12***	0.044***	0.076***
	(0.031)	(0.016)	(0.026)
Experiment (post)	0.11***	0.021	0.086***
	(0.034)	(0.015)	(0.031)
Week & duration FE	yes	yes	yes
Mean value dep. var.	0.95	0.23	0.73
Percent effect (during)	13	20	10
Percent effect (post)	11	9	12
N	73890	73890	73890

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by individual in parentheses.

Table 2: Regressions on weekly gym attendance (linear models): postexperiment effect w/o week 22 2017

	DiD model		
	(1)	(2)	(3)
	Visits	Class training	Free training
Experiment (during)	0.12***	0.044***	0.076***
	(0.031)	(0.016)	(0.026)
Experiment (post)	0.096***	0.021	0.075**
	(0.034)	(0.015)	(0.030)
Week & duration FE	yes	yes	yes
N	71427	71427	71427

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by individual in parentheses.

Table 3: Regressions on weekly gym attendance (Poisson models): postexperiment effect w/o week 22 2017

	DiD model		
	(1)	(2)	(3)
	Visits	Class training	Free training
Experiment (during)	0.13***	0.19***	0.11***
	(0.035)	(0.077)	(0.038)
Experiment (post)	0.11**	0.11	0.11**
	(0.046)	(0.092)	(0.052)
Week & duration FE	yes	yes	yes
Include pre-period	yes	yes	yes
N	64728	30450	60697

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by individual in parentheses. We report $\exp(\beta) - 1$, which is the percentage effect. Note that the numbers of observations in columns (1-3) vary because individuals with zero attendance in all time periods drop out in Poisson regressions with fixed effects.

Table 4: Regressions on weekly gym attendance (Poisson models): postexperiment effect until week 21 2017

	DiD model		
	(1)	(2)	(3)
	Visits	Class training	Free training
Experiment (during)	0.13***	0.19***	0.11***
	(0.035)	(0.077)	(0.038)
Experiment (post)	0.11**	0.10	0.11**
	(0.047)	(0.093)	(0.054)
Week & duration FE	yes	yes	yes
N	55600	25925	51750

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by individual in parentheses. We report $\exp(\beta) - 1$, which is the percentage effect. Note that the number of observations in columns (1-3) varies because individuals with zero attendance in all time periods drop out in Poisson regressions with fixed effects.

Table 5: Regressions on weekly gym attendance (Poisson models): excluding Christmas holidays

	Total visits		Classtraining		Freetraining	
	(1)	(2)	(3)	(4)	(5)	(6)
Experiment (during)	0.13***	0.11***	0.21***	0.24***	0.10***	0.076*
	(0.034)	(0.037)	(0.077)	(0.085)	(0.037)	(0.041)
Experiment (post)	0.12***	0.11**	0.13	0.15	0.12**	0.092*
	(0.046)	(0.049)	(0.093)	(0.10)	(0.052)	(0.055)
Week & duration FE	yes	yes	yes	yes	yes	yes
Exclude weeks:	52	51,52,01	52	51,52,01	52	51,52,01
N	64699	60156	30363	28107	60726	56295

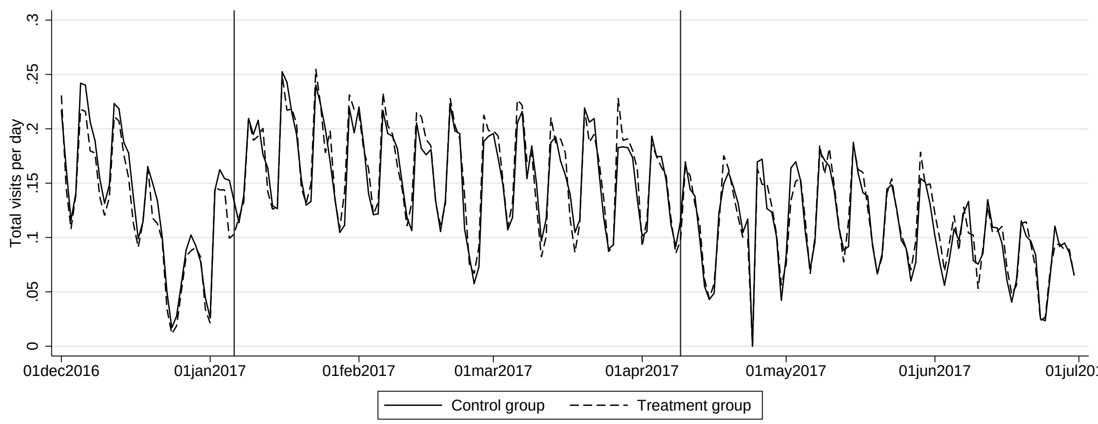
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by individual in parentheses. We report $\exp(\beta) - 1$, which is the percentage effect. Note that the number of observations varies because individuals with zero attendance in all time periods drop out in Poisson regressions with fixed effects. Columns (1), (3) and (5) exclude the last week of 2016 (week 52), in which attendance is very low. Columns (2), (4) and (6) in addition exclude week 51 (2016) and week 1 (2017) in which attendance is also below the trend. See Figure 4 in the paper for clarification.

Table 6: Regressions on weekly gym attendance (Poisson models): subsample with a longer pre-experiment period of October 1 - January 9

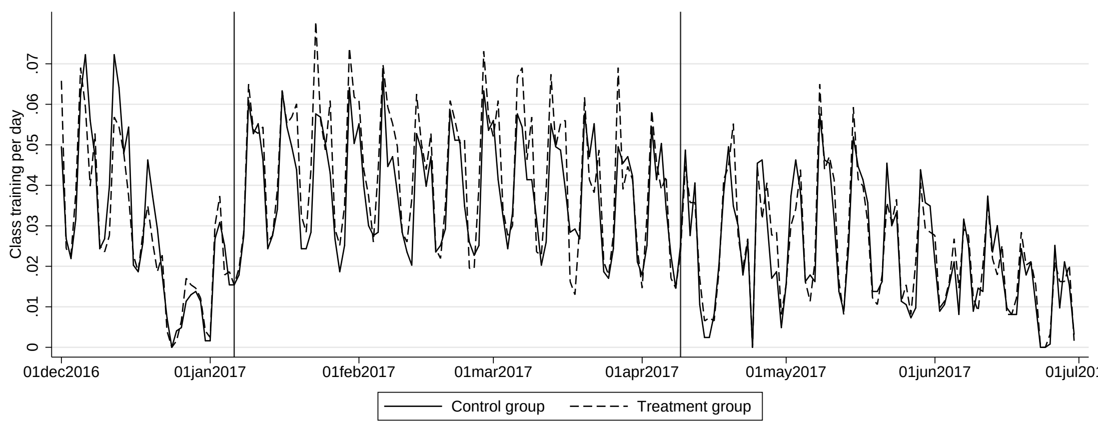
	DiD model		
	(1)	(2)	(3)
	Visits	Class training	Free training
Experiment (during)	0.051	0.20**	0.0037
	(0.038)	(0.088)	(0.041)
Experiment (post)	0.072	0.078	0.071
	(0.055)	(0.11)	(0.062)
Week & duration FE	yes	yes	yes
N	53430	27144	50778

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. We report $\exp(\beta) - 1$, which is the percentage effect. This longer pre-experiment period reduces the number of included individuals from 2,463 to 1,889. Note that the exact number of observations in the regressions varies because individuals with zero attendance in all time periods drop out in Poisson regressions with fixed effects.

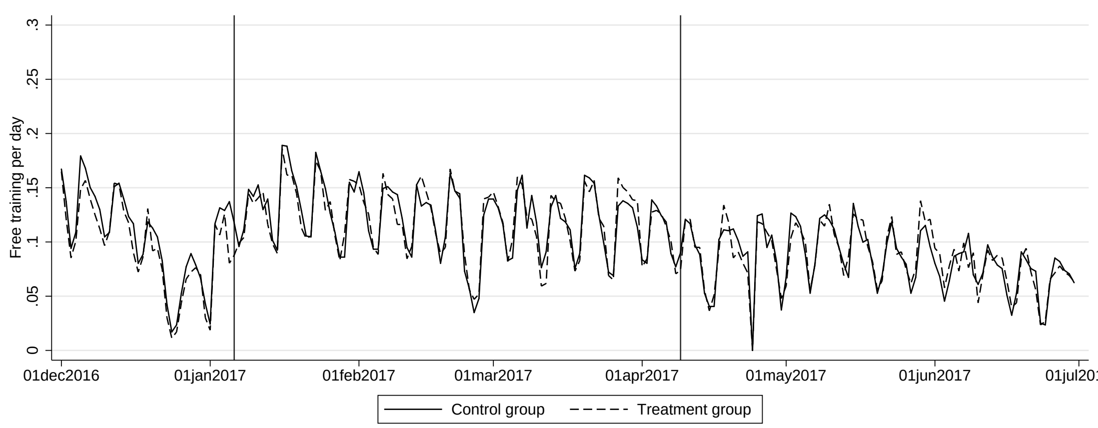
Figure 6: Gym attendance: daily data



(a) Total visits (daily)



(b) Class training (daily)



(c) Free training (daily)

Table 7: Regressions on daily gym attendance

	DiD model			Simple difference		
	(1) Visits	(2) Classtraining	(3) Freetraining	(4) Visits	(5) Classtraining	(6) Freetraining
Experiment (during)	0.014***	0.0056**	0.0088**	0.010*	0.0062*	0.0039
	(0.0042)	(0.0022)	(0.0035)	(0.0060)	(0.0033)	(0.0054)
Experiment (post)	0.012**	0.0040	0.0077	0.0065	0.0034	0.0031
	(0.0054)	(0.0026)	(0.0047)	(0.0065)	(0.0032)	(0.0059)
Recurring member				0.026***	0.0076**	0.018***
				(0.0063)	(0.0033)	(0.0057)
Male				0.024***	-0.040***	0.063***
				(0.0059)	(0.0033)	(0.0052)
Student				0.0019	-0.0013	0.0032
				(0.0072)	(0.0040)	(0.0064)
Daytimecard				-0.045***	-0.031***	-0.014
				(0.012)	(0.0063)	(0.011)
Sunny day				-0.0031*	0.00046	-0.0035**
				(0.0016)	(0.00081)	(0.0014)
Rainy day				-0.0082***	-0.0039***	-0.0043**
				(0.0020)	(0.0010)	(0.0018)
Week & duration FE	yes	yes	yes	yes	yes	yes
Day of week FE	yes	yes	yes	yes	yes	yes
Individual FE	yes	yes	yes	no	no	no
Include pre-period	yes	yes	yes	no	no	no
Mean value dep. var.	0.15	0.036	0.11	0.15	0.038	0.11
Percent effect	10	16	8	7	16	3
N	384228	384228	384228	259028	259028	259028

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by individual in parentheses. Seven age dummies and six gym location dummies are included but not reported in columns (4-6). A sunny day is a day with at least 5 hours of sunshine, and a rainy day is a day with at least 3 mm of rain.

Class popularity

We observe large variation across classes in terms of their popularity. While some classes are never fully booked and thus never have a waiting list, others are (almost) always fully booked. Arguably, the popularity of a class must have an impact on booking and cancellation decisions. For example, if a particular class is always fully booked, there is a need to book in advance because the class cannot be attended otherwise. If, by contrast, a class is never fully booked, there is no obvious need to book this class (unless the class's unpopularity is unknown to the individual).

As we discuss in Section 2.1, over 200 classes are scheduled each week, with 37 main class types (and further variation with respect to class length, focus, instructor, instruction language, etc.). While some classes are very popular, with all places being regularly taken, other classes are less popular (or are simply offered at higher capacity) and are rarely fully booked. Below, we describe how we split classes by popularity.

The first step is to define *a class*. Simply defining classes by their activity (e.g., Yoga, Pilates, etc.) and their scheduled time (e.g., Wednesday at 8 pm) will be too narrow. For example, if someone knows that '30 minutes Yoga' on Wednesday at 7 pm is very popular, she will most likely assume that '45 minutes Yoga' on Wednesday at 8 pm is also very popular (even if she never attended this particular class before). Thus, we need to define a class in a rather broad sense, such that similar classes at similar times fall into the same category. For this purpose, we first categorize each class into 37 main activities (these are the activity types, see the caption of Figure 1) in this appendix. Second, we assign each class to one of 14 time slots, defined by the day of the week interacted with 'daytime' or 'evening'. A particular class is thus defined as the activity type and the time slot. Examples of a class would be 'Yoga on Wednesday evening' or 'Football on Saturday daytime'. This approach also solves the problem that the exact scheduled time or name of a class might change over the course of one year. As a result, we observe 238 distinct classes, with between 5 and 5,617 bookings for each class. For 73 % of these classes, we observe at least 100 bookings during the 12 months.

Next, we categorize the 238 classes into two categories, based on how popular they are. We measure the popularity of a class by computing the share of bookings that result in a spot on the waiting list. A share of zero means that the class is never fully booked, which is the case for 30 % of all classes. A share of 0.05 means that one in 20 bookings ends up on the waiting list. So if, e.g., a class has 20 available spots, the class is typically fully booked. Of course, due to cancellations there is a reasonable chance of still being admitted to the class. However, we argue that an individual's perception of class popularity will be affected by (initially) ending up on the waiting list, and thus we opted for this particular measure. For 40 % of the classes, the share of bookings resulting in a spot on the waiting list is at least 0.05, while 22 % even have a share of more than 0.2 (indicating long waiting lists). We categorize the classes into unpopular (share less than 0.005) and popular (share more than 0.005).

Having categorized classes into popular and unpopular classes, we compare these two types of classes with respect to the number of bookings, cancellations and drop-in attendances in Table 8, columns (2)-(3), in this appendix. We find that, as expected, attendance through drop-in occurs more often in unpopular classes than in popular classes. However, even for unpopular classes, where waiting lists virtually never appear, more than half of all class participants (59 %) make a booking. This finding is surprising, given that there is no actual need to make a booking for unpopular classes. This finding also holds when splitting the sample into 'experienced' and 'inexperienced' attenders, see Table 9 in this appendix. The former have attended the same class (type) at least five times. While we expect most gym members to have reasonable knowledge about the popularity of a class, such experienced visitors certainly know that a booking is never necessary for unpopular classes. We take this finding as evidence that many individuals value bookings as a kind of commitment device to overcome time-inconsistent preferences, i.e., present bias (O'Donoghue and Rabin, 1999a,b, 2001). In the context of exercising, present bias implies that individuals continuously make plans to exercise in the future, but decide to postpone these plans once the day of the planned activity approaches.

The intuition for bookings serving as a commitment device goes as follows: some individuals benefit from signing up to an unpopular class (they will only do so if they are aware of their time inconsistency, i.e., if they suffer from sophisticated time inconsistency), because this raises their cost of not attending. Since canceling or missing a booking is costly in terms of the hassle of cancellation or payment of a fee if one does not show up for the booked class, attendance becomes more likely. Bookings are thus a form of a commitment contract, which, in contrast to commitment contracts in other contexts, are not introduced as part of a field experiment but occur naturally in our setting. The repeated take-up of these contracts also implies that the decision is less likely to be an error, as may be the case for one-off decisions. Our findings complement recent studies that demonstrate the demand for and effectiveness of commitment contracts at the stage of buying a gym membership (e.g., Goldhaber-Fiebert et al., 2010; Bhattacharya et al., 2015; Royer et al., 2015).

Furthermore, the share of no-shows is the same across all types of classes (see Table 8 in this appendix). This finding suggests that limited attention is a common phenomenon and does not depend on the type of class.

Table 8: Bookings, cancellations and drop-in attendance by class popularity

	All classes	Unpopular classes (< 0.5% waiting list)	Popular classes (> 0.5% waiting list)
Number of bookings	43,953	5,876	38,077
Of those (in shares):			
Attendance	0.52	0.56	0.51
Cancellation	0.42	0.39	0.42
No-shows	0.054	0.052	0.054
On final waiting list	0.008	0.000	0.009
Number of drop-in attendance	6,114	2,235	3,879
As share of all attendance	0.21	0.41	0.17
Number of classes	235	95	140

The categorization of classes by popularity is made based on the share of all bookings that goes to the waiting list (which is, due to cancellations, much larger than the share that eventually remains on the list when the class occurs). The 'Cancellation' share includes cancellations by people on the waiting list.

Table 9: Booking and canceling by class popularity: experienced attenders that have attended at least 5 times before

	All classes	Unpopular classes (< 0.5 % waiting list)	Popular classes (> 0.5 % waiting list)
Number of bookings	15,017	1,395	13,622
Of those (in shares):			
Attendance	0.58	0.67	0.57
Cancellation	0.37	0.29	0.38
No-shows	0.04	0.04	0.04
On final waiting list	0.01	0.00	0.01
Number of drop-in attendance	2,075	952	1,123
As share of all attendance	0.19	0.50	0.13
Number of classes	169	50	119

The categorization of classes by popularity is made based on the share of all bookings that goes to the waiting list (which is, due to cancellations, much larger than the share that eventually remains on the list when the class occurs). The 'Cancellation' share includes cancellations by people on the waiting list.

Additional references (not contained in the paper)

Goldhaber-Fiebert, J. D., Blumenkranz, E., and Garber, A. M. (2010). Committing to exercise: Contract design for virtuous habit formation. *NBER Working Paper*, No. 16624.

Bhattacharya, J., Garber, A. M., and Goldhaber-Fiebert, J. D. (2015). Nudges in exercise commitment contracts: A randomized trial. *NBER Working Paper*, No. 21406.