# Longitudinal surveys after medication dispensing: real-world data from an at-home smart hub

Benjamin Alexander Ogorek, Thomas Patrick Rhoads, Erica Ann Smith

Submitted to: Journal of Medical Internet Research
on: May 10, 2024

# *Table of Contents*

# Longitudinal surveys after medication dispensing: real-world data from an at-home smart hub

Benjamin Alexander Ogorek[1] PhD; Thomas Patrick Rhoads[2] MBA; Erica Ann Smith[2] PhD

[1]Spencer Health Solutions Inc Morrisville US

**Corresponding Author:**
Benjamin Alexander Ogorek PhD

## *Abstract*

**Background:** Real-World Data (RWD) has gained strategic, financial, and regulatory importance in the pharmaceutical industry during both development and commercialization. However, the actual RWD submitted to regulatory bodies often fails to be considered primary evidence. Fit for purpose longitudinal survey methods allow researchers to be targeted in their data acquisition but face the challenges of panel recruitment, panel attrition, and low data quality from an incentivized population. An untapped platform for administering longitudinal surveys to patients is the digital medication dispensing device. One such system is an in-home smart hub called "spencer."

**Objective:** We evaluated whether the spencer stand-alone medication dispensing smart hub could be a primary generator of longitudinal patient reported outcomes. Using the spencer platform, we measured platform persistency, survey compliance, and the psychometric properties of the survey responses.

**Methods:** We analyzed 4,138 patients on the spencer smart hub platform during their first two years of tenure. A discrete survival framework was used to estimate platform persistency. Survey opt-out and response rates were computed by patient tenure. For accessing reliability and validity, we examined a spencer question on reported falls called "Q_FALL." For reliability, we looked at the within-patient correlation between tenure years one and two mean Q_FALL and computed transition rates between intervals, assuming persistence of vigor among healthy patients. For evidence of validity, we measured the degree of association between Q_FALL and known factors influencing fall risk, including age, biological sex, quality of life, physical and emotional health, and use of SSRIs or SNRIs.

**Results:** For the spencer platform, 51.0% of patients were retained beyond two years (24 thirty-day periods). For 1,832 patients retained past one year, 82.0% kept their surveys enabled through the 12th month of tenure with an aggregate response rate of 84.1% in the period. Two thirds of patients had near-perfect response rates for the entire year while one third answered less frequently over time. For 234 patients with >= 5 Q_FALL responses in the first two years of tenure, the within patient Pearson correlation was 0.723. In the rare falls category, 84% in year one remained in the same category in year two. Q_FALL did not show the expected relationship with sex (P=.66) or age (P=.76) but had significant positive relationships (all having P<0.001) with SSRI / SNRI usage, quality of life, depressive symptoms, physical health, disability, and trips to the emergency room.

**Conclusions:** Longitudinal surveys administered on a medication dispensing smart hub can generate robust, fit for purpose RWD. Patients were persistent to the spencer platform for years and continued to answer survey questions at high rates. Response patterns showed evidence that reliable and valid measures of important health constructs are possible.

(JMIR Preprints 10/05/2024:60438)
DOI: https://doi.org/10.2196/preprints.60438

## Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

# Original Manuscript

# Longitudinal surveys after medication dispensing: real-world data from an at-home smart hub


## Abstract

**Background:** Real-World Data (RWD) has gained strategic, financial, and regulatory importance in the pharmaceutical industry during both development and commercialization. However, the actual RWD submitted to regulatory bodies often fails to be considered primary evidence. Fit for purpose longitudinal survey methods allow researchers to be targeted in their data acquisition but face the challenges of panel recruitment, panel attrition, and low data quality from an incentivized population. An untapped platform for administering longitudinal surveys to patients is the digital medication dispensing device. One such system is an in-home smart hub called “spencer.”

**Objective:** We evaluated whether the spencer stand-alone medication dispensing smart hub could be a primary generator of longitudinal patient reported outcomes. Using the spencer platform, we measured platform persistency, survey compliance, and the psychometric properties of the survey responses.

**Methods:** We analyzed 4,138 patients on the spencer smart hub platform during their first two years of tenure. A discrete survival framework was used to estimate platform persistency. Survey opt-out and response rates were computed by patient tenure. For accessing reliability and validity, we examined a spencer question on reported falls called “Q_FALL.” For reliability, we looked at the within-patient correlation between tenure years one and two mean Q_FALL and computed transition rates between intervals, assuming persistence of vigor among healthy patients. For evidence of validity, we measured the degree of association between Q_FALL and known factors influencing fall risk, including age, biological sex, quality of life, physical and emotional health, and use of SSRIs or SNRIs.

**Results:** For the spencer platform, 51.0% of patients were retained beyond two years (24 thirty-day periods). For 1,832 patients retained past one year, 82.0% kept their surveys enabled through the 12$^{th}$ month of tenure with an aggregate response rate of 84.1% in the period. Two thirds of patients had near-perfect response rates for the entire year while one third answered less frequently over time. For 234 patients with >= 5 Q_FALL responses in the first two years of tenure, the within patient Pearson correlation was 0.723. In the rare falls category, 84% in year one remained in the same category in year two. Q_FALL did not show the expected relationship with sex ($P$=.66) or age ($P$=.76) but had significant positive relationships (all having $P$<0.001) with SSRI / SNRI usage, quality of life, depressive symptoms, physical health, disability, and trips to the emergency room.

**Conclusions:** Longitudinal surveys administered on a medication dispensing smart hub can generate robust, fit for purpose RWD. Patients were persistent to the spencer platform for years and continued to answer survey questions at high rates. Response patterns showed evidence that reliable and valid measures of important health constructs are possible.


**Keywords:** Real-world data; real-world evidence; patient-reported outcomes; RWD; RWE; longitudinal studies; survey methods

## Introduction

## Background

In the early 2000s, the use of Real-World Data (RWD) by the pharmaceutical industry was rooted in the domain of safety and surveillance.[1] The passing of the Cures Act in 2016 by the United States

Congress and subsequent initiation of a Real-World Evidence (RWE) Program[2] by the US Food and Drug Administration (FDA) in 2018 elevated RWD as a tool to support new indications for already approved drugs.[34] The RWD and RWE trend is global; the European Medicines Agency (EMA) published their RWE framework in 2023,[5] and Canada's Drug and Health Technology Agency published their "guidance" document the same year.[6]

RWE has been submitted and considered as part of recent FDA applications. One study found that out of 136 FDA application approvals between 2019 and 2021, 116 used RWE either to support therapeutic context, safety, or effectiveness. Out of the 88 studies using RWD to support safety or effectiveness, the FDA addressed all but 10 submissions, finding that 57 offered "supportive evidence," 8 offered "substantial or primary evidence," and 13 were "not adequate for decision making." Feedback from the FDA on the submitted evidence included concerns about missing data, small sample sizes, bias, and even the "premise" of the study.[7]

Examples of RWD are claims records, electronic health records, registries, and patient-generated data, including patient reported outcomes (PROs). PROs have been described by the FDA as "a questionnaire plus the information and documentation that support its use,"[8] but an important distinction from other types of patient-generated health data is that PROs are generated by the patient. "In these instances, the patient is acting as a proxy research observer, observing symptoms and events over time and across a spectrum of real world environments and circumstances."[9] While electronic health records are "approaching the genomic scale in volume and complexity"[10] patient reported outcomes will be relatively limited in volume because they rely on the patient to actively contribute responses.

Longitudinal surveys, where respondents are interviewed at two or more points in time, are easier to execute thanks to the proliferation of internet-based surveying.[11] These allow for the analysis of within-unit change as well as aggregations over time,[12] and offer greater "causal leverage" than cross-sectional surveys. This is especially true when there is an intervention between the repeated surveys.[11] Online panels, or "[pools] of registered persons who have agreed to take part in online studies on a regular basis," also rose in attractiveness with the proliferation of the internet.[13] For example, the Patient-Reported Outcomes Measurement Information System (PROMIS), which contains global health survey items, has seen successfully used with convenience panels designed to represent the general population.[14]

Despite these advantages, longitudinal studies based on online panels face multiple difficulties. On the recruitment side, the composition of online convenience panels is known to differ from the underlying population and response rates are often lower than 10%. Though quota sampling can mitigate these shortcomings, problems with the quality of internet panel data include "false answers, careless responses, giving the same answer repeatedly, getting multiple surveys from the same respondent, and panelists being members of multiple panels."[14] Quality checks are possible but can greatly reduce the available sample size. In one study, an attempt to use Amazon Mechanical Turk to build a diabetes panel failed after only 6% (n=13) were deemed eligible for future survey research.[15]

On top of recruitment and data quality issues, longitudinal studies face "panel attrition," where patients in earlier waves cease to respond in later waves.[11] This is especially problematic for studies featuring mobile apps. For eight remote digital studies conducted between 2014-2019, consisting of more than 100,000 participants, researchers found that more than half of all participants discontinued their participation within the first week of the study.[16] In a study of smartphone app usage to improve oral anticoagulation adherence in patients with atrial fibrillation, a retention rate of 27% at six months was reported.[17]

The advent of in-home patient engagement technology provides new options to survey patients longitudinally as the recruited patient can double as an online panel. On example is a medication dispensing smart hub called “spencer” that combines medication dispensing with collection of both survey data and other patient-generated data. Patel et al reported 98% medication adherence (defined as a dispense within two hours of the scheduled time) with the spencer platform, along with reduced caregiver burden.[18] Previous studies have used smartphones to measure or improve medication adherence[19,20] with modest results. For instance, an SMS-based intervention increased proportion of days covered from 79% to 83%.

The spencer platform can generate RWD directly from the home via survey questions displayed on the touchscreen directly following an on-time medication dispense. Using the spencer platform alone, this method has generated over three million longitudinal responses from more than 4,000 unique patients to health-related questions from a chronic care population.

## Objective

The goal of this research was to evaluate the feasibility of using the spencer medication dispensing device as a primary generator of longitudinal patient reported outcomes. This study sought to quantify patient persistence to the spencer platform, survey response rates, and the validity and reliability of these measures.

## Methods

## Recruitment

A Canadian value-based care organization has recruited patients into a spencer medication therapy management program since 2018. The spencer units were shipped to the patients’ homes upon recruitment. Once set up, they display both clock time and the scheduled time of the next medication dispense via a touch screen display. Refills containing medication strips packaged by a pharmacy or clinical packager were regularly shipped to the patients’ homes and were inserted into the top of the unit via an electronically controlled door. At scheduled times, the spencer unit alerted the patient through sound and light from the built-in touchscreen display notifying them of the need to take medication. When the patient pressed a button on the touchscreen, the unit dispensed one or more medication pouches. Afterwards, a question was presented to patients who have not explicitly opted out. If the patient wished to answer, two button presses were necessary, one for the response selection and one for the confirmation.

The inclusion criteria for this study were patients onboarded to the spencer platform who had medication scheduled during the observation window of 2021-06-03 to 2024-02-14. Patients were considered “onboarded” after dispensing medications with spencer (for the purpose of this paper, “dispensing medications” means dispensing multi-dose packs containing medication).

Date of birth and biological sex were entered using an online portal when patients are recruited, and these fields were retrieved from the application database on 2024-03-04. Dates of birth that were within 2 years of the database added date were replaced with missing values as these are indicative of the current date being recorded as the birth date. We then computed the age as the difference between the first dispense date and date of birth. Information about the drugs scheduled for these

patients entered the database when medication refills were created by the pharmacies.

## Statistical Analysis

### Platform Persistency

Patients leave the spencer platform for multiple reasons, including life transitions to higher care services and a patient's natural death. For estimating platform persistency, we used the discrete survival analysis framework described by Allison.[21] The discrete periods start on the first day a patient is scheduled, are 30 days in length, and the attrition event occurs when a patient is not scheduled during an entire 30-day period. When the patient is scheduled in a later period after previously meeting the definition for an attrition event, the attrition flag is reset for all previous periods.

### Survey Response Rates

We studied survey response rate patterns from patients who stayed on platform past 14 thirty-day periods (a year and additional time for the first dispense to occur) and where the analysis window overlapped with the first year of tenure.

Other than attrition from the spencer platform itself, there are two other mechanisms of loss of survey data. The first is when a patient requests for their surveys to be turned off following medication dispensing. The second is when patients do not enter a response when a question is displayed, either by leaving the confirmation button unpressed or ignoring the screen entirely following the medication dispense. To investigate the rates of survey data loss due to these two mechanisms, we constructed time series plots of the percent of patients that were still receiving surveys (i.e., had not requested them to be turned off) and the percent of displayed surveys that were responded to. Both series were plotted against tenure during the first year on platform.

For the survey response rate data, we knew from prior analyses that some patients consistently respond and wished to study this subgroup. We used a threshold of 90 percent in period 12, slightly above the averages observed in weekly reporting, to create two groups. We then plotted response rate time series for both.

## Psychometric Analysis

### Reliability

Reliability is the consistency of a measure. The spencer longitudinal survey platform would produce unreliable measures if, for instance, patients randomly chose responses after their dispenses. Since an entire platform is too broad for a reliability query, we looked for evidence that a single question asked through the spencer produced a reliable measure.

Inspired by the "Falls Efficacy Scale-International" (FES-I), a reliable measure of fear of falling (itself related to actual falls),[22-23] we chose a standard spencer question that asks about recent falls. Hereafter referred to as “Q_FALL,” the question prompt is “*Have you experienced a fall in the past month?”* and the response options are *“No,” “Not Sure,” “Yes”* (a one-letter variation in capitalization occurring after September 2022, where “*Not sure”* was replaced with *“Not Sure”*). These values were value coded as 1, 0, and -1, respectively, and we averaged these values over time.

There exist multiple reliability concepts, including “internal consistency reliability” and “interrater reliability,”[24] and we focused on test-retest reliability for Q_FALL. Test-retest reliability measures the similarity between measurements taken at multiple points in time when the underlying construct has not meaningfully changed. FES-I measurements have previously been compared and observed to be “moderately stable over time” with Pearson correlations ranging from 0.66 to 0.83 for measurements taken up to a year apart.[25] We examined the Pearson correlation coefficient between the mean Q_FALL scores for patients in tenure year one vs tenure year two to observe whether its value was near those seen with serial FES-I measurements.

To compute the Q_FALL based measures over two tenure years, we limited attention to a subset of 234 patients from the persistency analysis who stayed on platform past 26 thirty-day periods (two years plus extra time for the first dispense) and answered Q_FALL at least 5 times in both the first and second year of tenure. Averages of integer coded values were computed within patient for each of the two tenure years. We then compared the Pearson’s correlation coefficient between years one and two to the range of 0.66 to 0.83.[25]

Additionally, we hypothesized that patients who never or rarely indicated a recent fall would tend to do so in both years, as balance, stability, and vigor would likely persist across two years of life. We examined the frequency of transitions to and from the ranges mean Q_FALL scores of [-1, -0.5), [-0.5, 0.5) and [0.5, 1.0] in tenure years one and two, expecting that patients with a mean year one Q_FALL within [0.5, 1.0] to be the most likely to remain in that interval in year two.

### *Validity*

Validity is “the extent to which the scores from a measure represent the variable they are intended to.”[24] This translates into whether responses to Q_FALL have a relation to actual patient falls. To access the convergent validity of mean Q_FALL, we compared it known fall risk factors of increased age, biological sex, previous fall frequency, low quality of life, and depressive symptoms, physical impairment, and medication use.[22,25-26] Regarding medication use, many patients in the population are prescribed SSRIs and SNRIs, medications which are associated with falls in the elderly.[27,28] Thus, we considered an SSRI and SNRI prescription as a risk factor.

We undertook a series of analyses to detect associations that would constitute evidence of convergent criterion validity. We used the 234 patients from the reliability analysis, as these patients persisted for at least two years on the spencer platform and had at least 5 Q_FALL responses in each of their first two years of tenure. There were three categories of factors to test for associations with Q_FALL. The first is the autoregressive predictor, using previous falls to predict future falls. Though this could fall under predictive validity, we studied within-patient stability in the reliability section and did not consider it again as part of validity. Second, there are the true covariates such as age and sex which are known before any responses are received. We also included SSRI and SNRI medication usage in this category since the use of medication classes is generally stable within patient. Finally, there are contemporaneous responses from theoretically related questions received from spencer, alternate outcome measures, that should correlate with Q_FALL.

For the analysis of the relationship between Q_FALL and covariates, age was allowed to vary within patient for this analysis and was calculated as the number of years from the recorded date of birth to the date of each response. Biological sex was queried from the database on March 4$^{th}$, 2024. To accommodate the repeated measures received from each patient, we used a generalized estimating equation (GEE) approach to model the relationship between the coded value of Q_FALL with an

exchangeable working correlation structure (via the *geepack* package in R). The resulting standard errors are robust to both the choice of working correlation structure and non-normality of the response.[29]

For analysis of the relationship between Q_FALL and contemporaneous outcomes, we used the robust Kendall's tau measure to quantify the strength of association between pairs of within-patient means. Kandall's tau has an accompanying two-sided nonparametric test for testing the null hypothesis of zero association, which we relied on due to the limited number of response categories and different response counts between patients. The spencer questions comprising the list of contemporaneous outcomes are in Table 1.

Table 1. Standard spencer questions relating to known risk factors of falling.

| Question Text | Possible Responses | Values coded | Construct[a] |
|---|---|---|---|
| | | | |
| Rate your recent quality of life | Excellent \| Very Good \| Good \| Fair \| Poor | 5 \| 4 \| 3 \| 2 \| 1 | Quality of life |
| How is your emotional health today? | Excellent \| Very Good \| Good \| Fair \| Poor[b] | 5 \| 4 \| 3 \| 2 \| 1[b] | Depression |
| How would you rate your physical health today[sic] | Excellent \| Very Good \| Good \| Fair \| Poor | 5 \| 4 \| 3 \| 2 \| 1 | Physical health |
| Rate your ability to perform activities today | Excellent \| Very Good \| Good \| Fair \| Poor | 5 \| 4 \| 3 \| 2 \| 1 | Ability / Disability |
| Are you able to accomplish what you have planned today? | Completely \| Mostly \| Moderately \| A little \| Not at all | 5 \| 4 \| 3 \| 2 \| 1 | Ability / Disability |
| Hospital, ER or Urgent Care in the past month? | No \| Not Sure \| Yes | 1 \| 0 \| -1 | ER visits from falls |

[a] Construct is based on face validity of the spencer standard questions.
[b] This question has been in rotation for multiple years, but in September 2022, the number of responses changed from three *("Poor", "Good", "Excellent")* to five *("Poor", "Fair", "Good", "Very Good", "Excellent"*). We coded the three-response set as 1, 3 and 5, and the five-response set as 1, 2, 3, 4, 5, respectively.

## Results

## Patient Population

Out of the 4,138 patients, all but 5 patients had a recorded sex; the distribution was 2,552 (61.7%) female and 1,581 male (38.3%). There were 15 patients for which the date of birth was within 2 years of the data entry date; these were discarded, leaving 4,123 patients with recorded ages. The median (mean) age was 55 (54.4) years, the 1st and 3rd quartiles were 39 and 70, and the range was from 5 to 104 years. All but two of the 4,138 patients had a postal code that was successfully linked

to a specific region. Of these, most patients were from Canada, namely Ontario (2,083), British Columbia (1,149), or Saskatchewan (485). Within the US, the patients were primarily concentrated in Tennessee (221), Missouri (117), California (32) and Ohio (24).

This patient population was characterized by multiple chronic conditions treated by numerous concurrent medications. Patient were on average scheduled to take 10 different drugs per day. Of the 2,805 unique compounds observed in this population, 70% were mapped to an Anatomical Therapeutic Chemical (ATC) Classification System Level 2 code, with Vitamins & Supplements as a modified category that was inclusive than the ATC category of Vitamins. Table 2 contains the 20 most frequently observed mapped ATC Level 2 subgroups taken by these patients. The high frequency of psychoanaleptics indicates a high frequency of treatment for mental health concerns in this population. Vitamin D was the most common single compound dispensed. Drugs for treating cardiovascular conditions were also very common.

Table 2. The most frequently observed Anatomical Therapeutic Chemical (ATC) System Level 2 codes among the 4,138 patients studied.

| Rank | ATC Level 2 Subgroup | Number of Patients |
|---|---|---|
| | | |
| 1 | Psychoanaleptics | 3,040 |
| 2 | Vitamins & Supplements | 1,917 |
| 3 | Lipid Modifying Agents | 1,848 |
| 4 | Drugs for Acid Related Disorders | 1,788 |
| 5 | Antiepileptics | 1,717 |
| 6 | Agents Acting on the Renin-Angiotensin System | 1,664 |
| 7 | Psycholeptics | 1,503 |
| 8 | Drugs Used in Diabetes | 1,317 |
| 9 | Beta Blocking Agents | 1,107 |
| 10 | Anti-Thrombotic Agents | 1,050 |
| 11 | Calcium Channel Blockers | 893 |
| 12 | Diuretics | 822 |
| 13 | Thyroid Therapy | 741 |
| 14 | Urologicals | 592 |
| 15 | Analgesics | 543 |
| 16 | Anti-Inflammatory and Antirheumatic Products | 378 |
| 17 | Anti-Hypertensives | 345 |
| 18 | Anti-Parkinson Drugs | 309 |
| 19 | Drugs for Constipation | 295 |
| 20 | Antihistamines for Systemic Use | 289 |

## Platform Persistency

The persistency curve estimated from the 4,138 patients over 24 thirty-day periods is shown in Figure 1. The percentage of patients scheduled beyond the 12$^{th}$ period was 68.3%, which drops to 51.0% beyond the 24$^{th}$ thirty-day period. The rate of attrition was slightly higher in the first six months, settling down to a roughly constant rate near 2% per period.

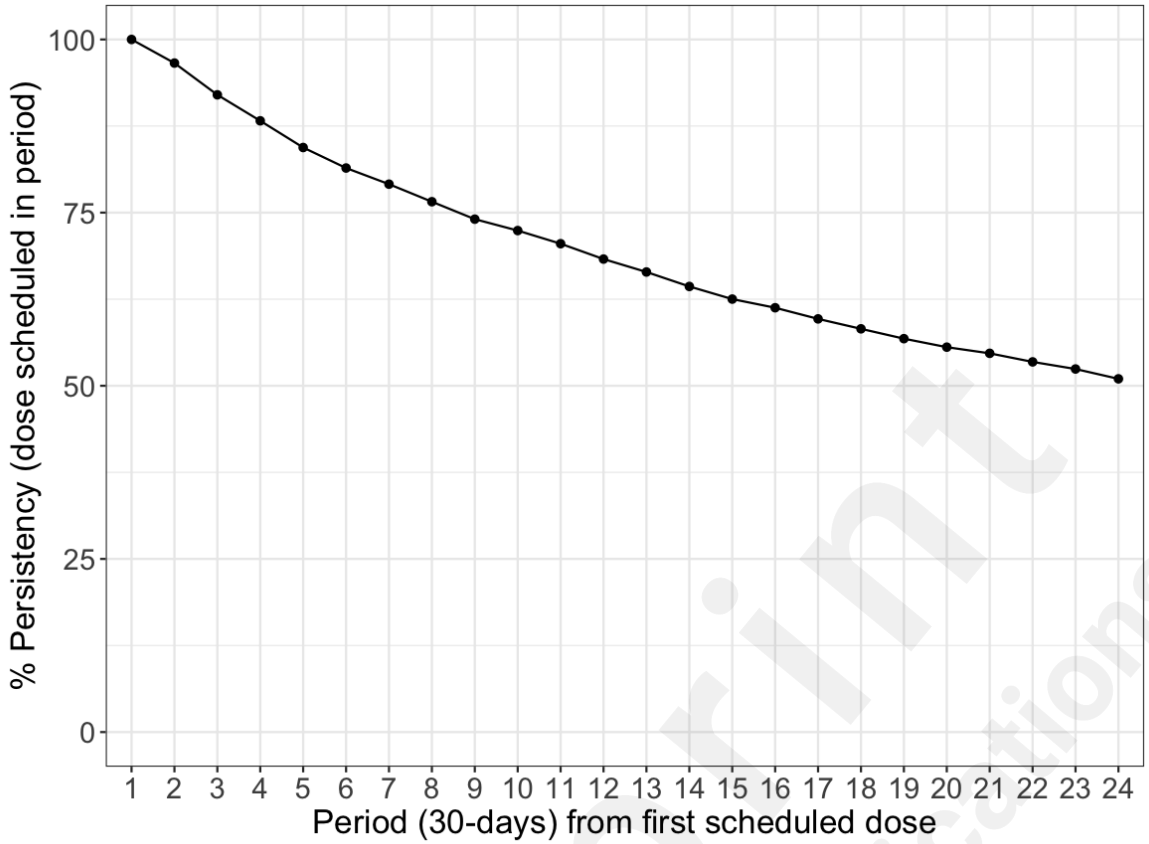



Figure 1: spencer platform persistency over the first two years of patient tenure.

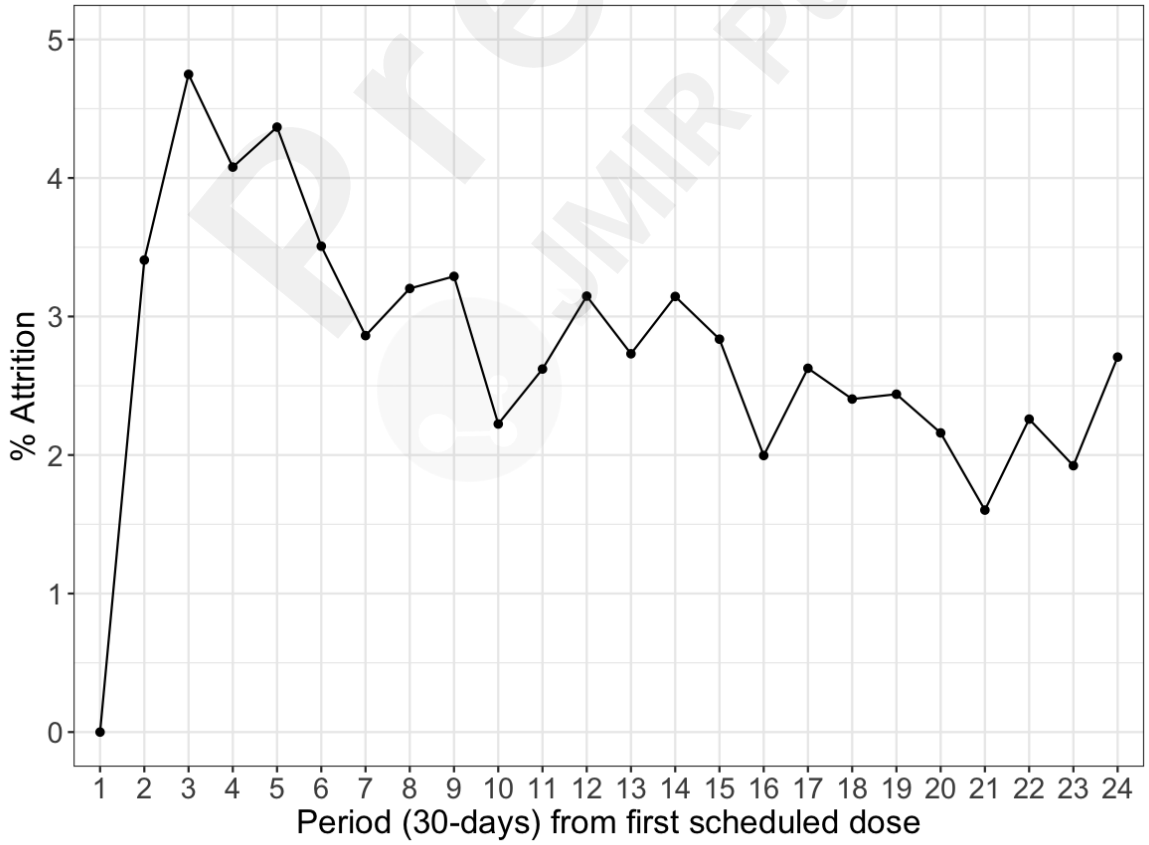


Figure 2. spencer platform attrition rate over the first two years of patient tenure.

## Survey Response Rates

Among the 1,832 patients who stayed on the spencer platform past 14 periods, the discontinuation rates were between 0.6% and 3.2% in the first 12 periods. They were generally decreasing, with the highest rate in second period (3.2%) and the lowest rate in the eleventh and twelfth periods (0.6% and 0.7%, respectively).

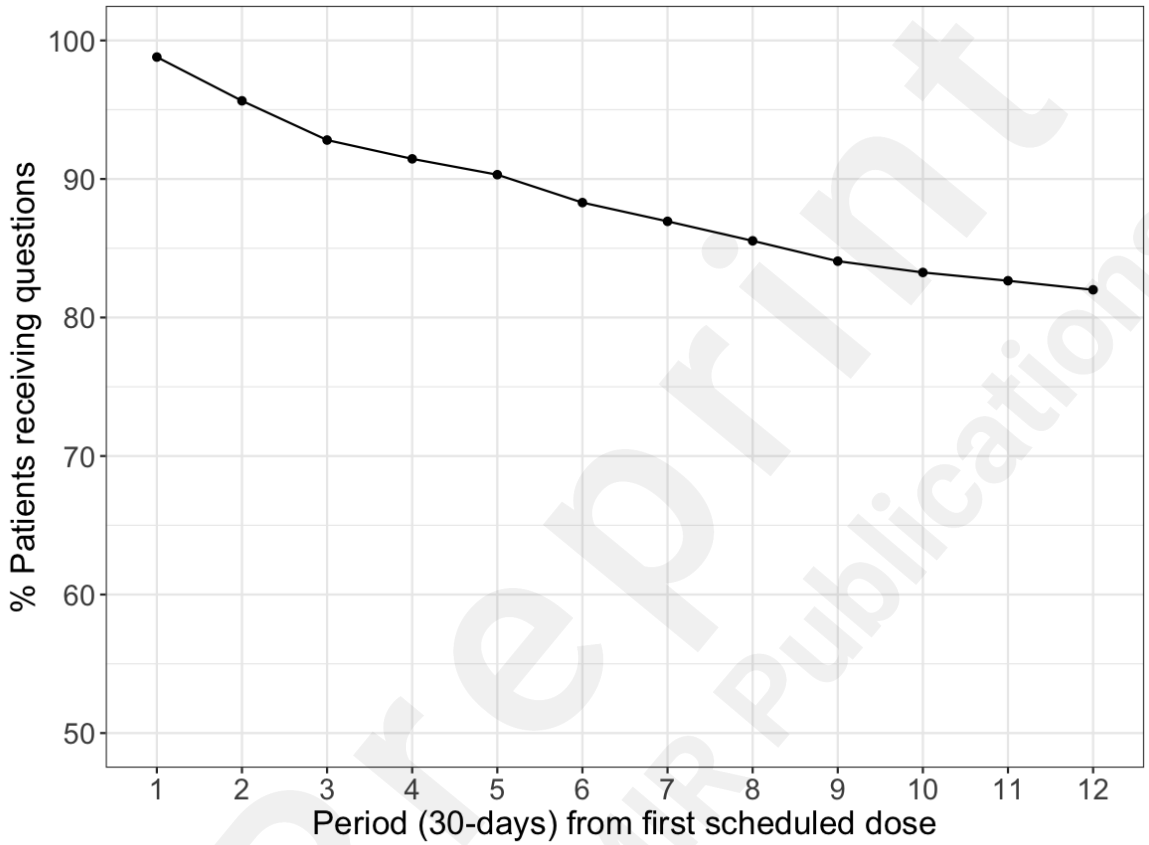



Figure 3. Proportion of patients who continue to receive surveys during their first year of tenure on the spencer platform.

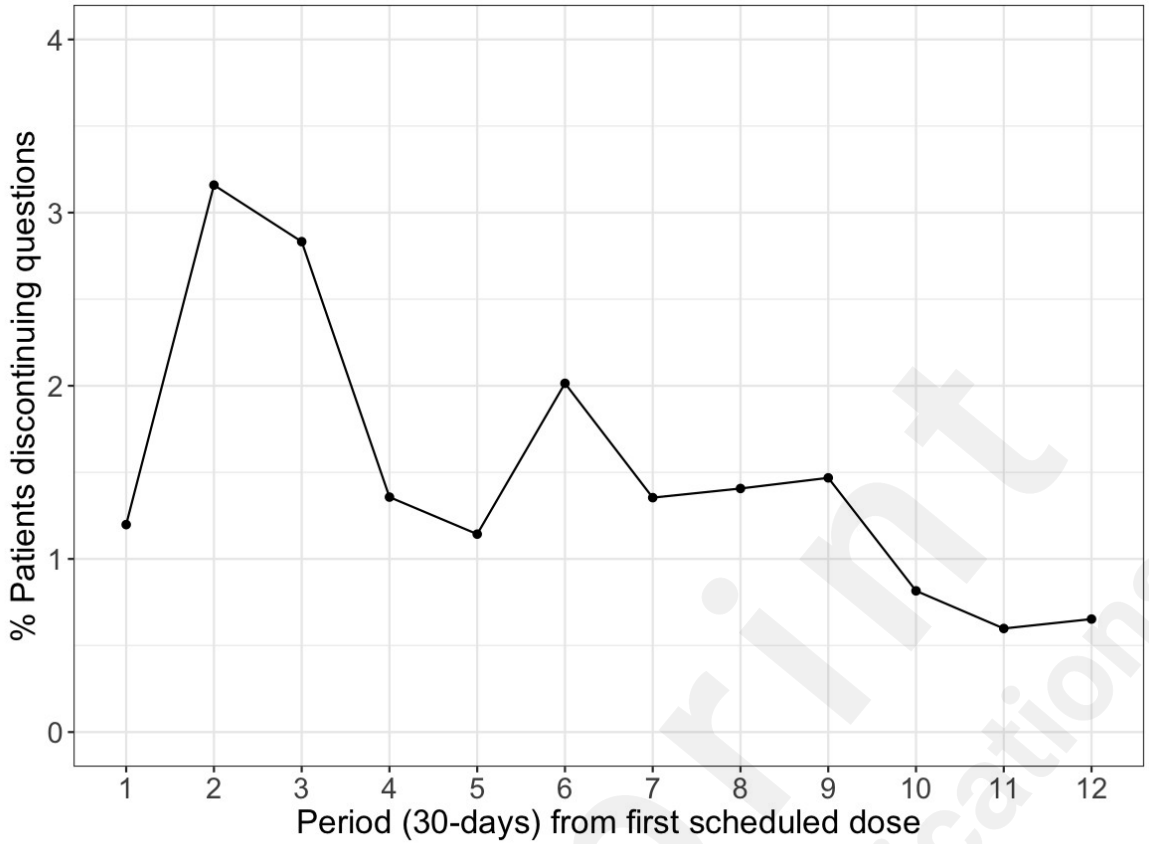



Figure 4. Discontinuation rate for survey questions during the patient's first year of tenure on the spencer platform.

For the survey questions themselves, the response rate generally declined from 95.6% in the first period to 84.1% in the 12$^{th}$ period (see Figure 3). After segmenting the population based on whether the period 12 response rate (RR) was above or below 90%, we found that two thirds (67.9%) of the population had nearly perfect response rates, ranging from 97.2% to 98.8% throughout the entire first year of tenure, whereas the remaining one third experienced decreasing response rates from 91.7% in period 1 to 53.1% in period 12 (see Figure 4).

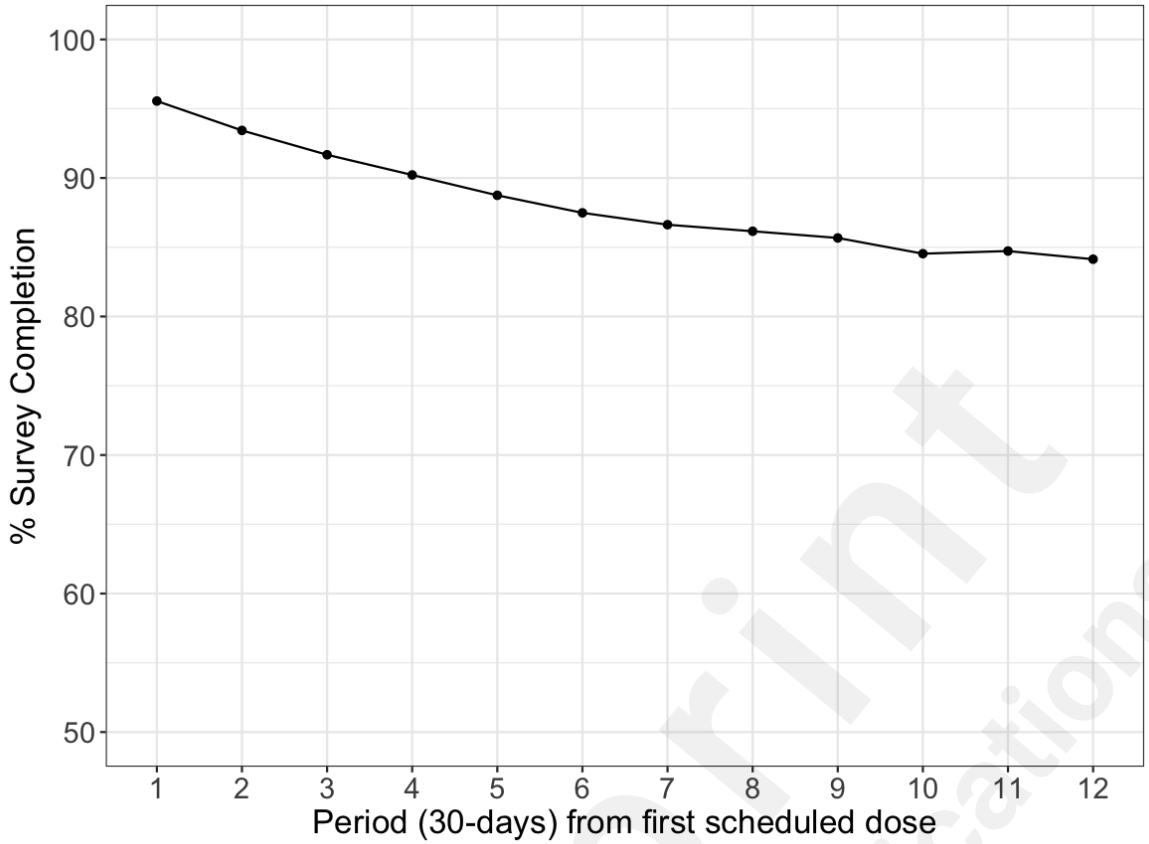



Figure 5. Proportion of surveys that are responded to during the first year of tenure among patients who stay on platform.

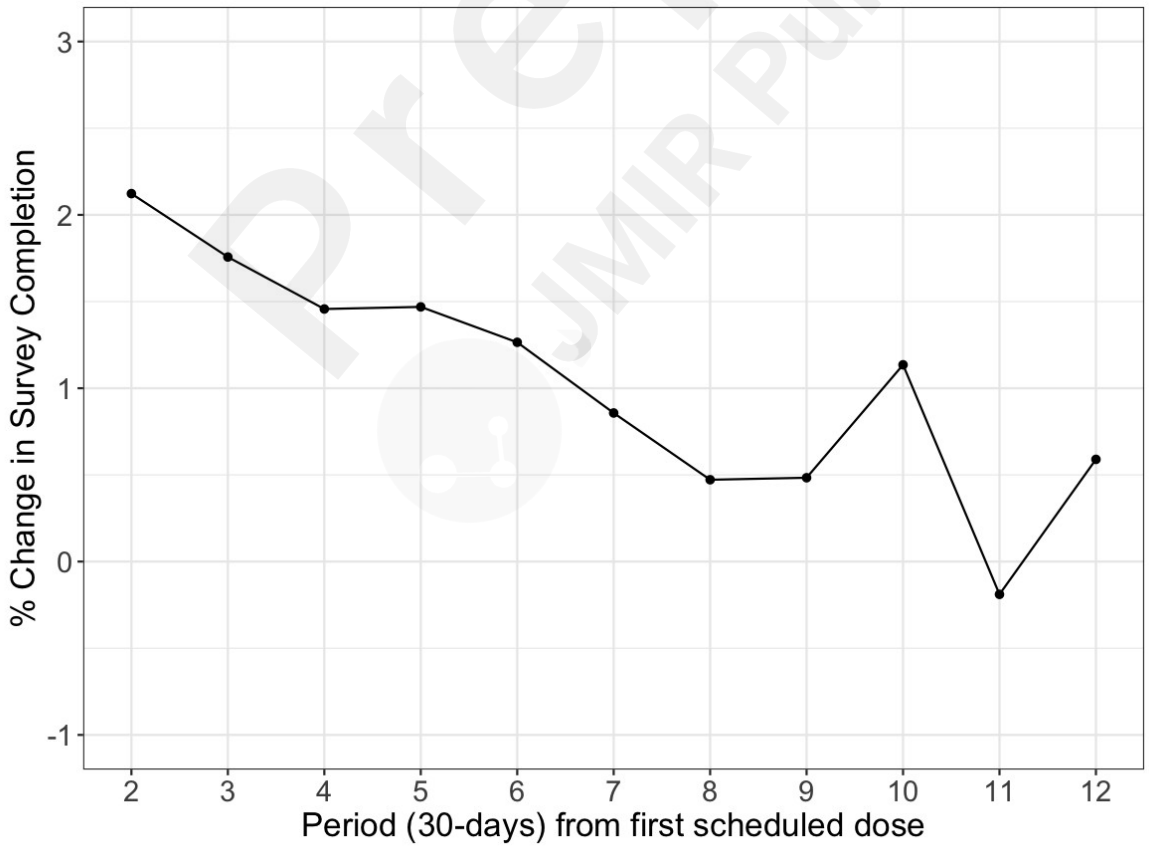


Figure 6. Change in survey completion rate over the first year of tenure.

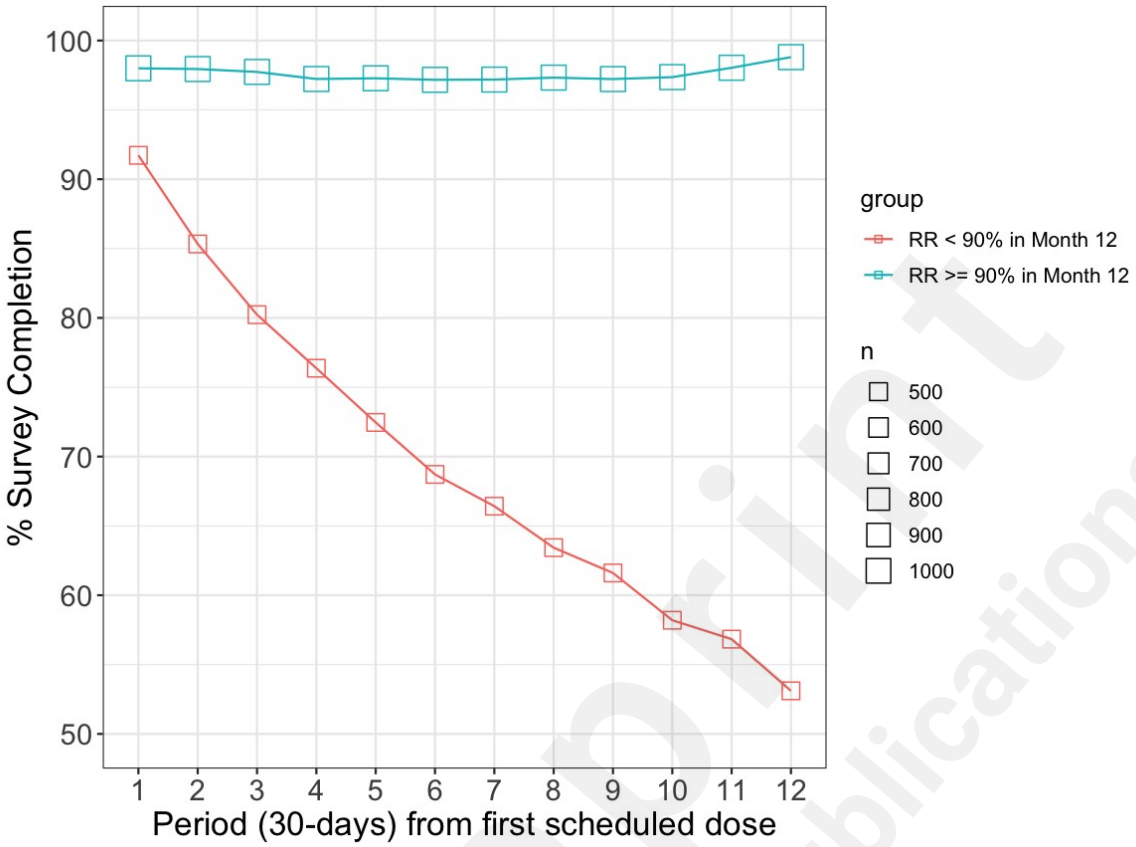



Figure 7. Proportion of surveys that are completed, grouped by whether the survey response rate (RR) was >=90% or < 90% in period 12.

## Reliability

Among the 234 patients with at least 5 responses to Q_FALL in each of the first two years of tenure with the spencer platform, the Pearson correlation between the year one and two means was 0.723, which was within the range of values seen in FES-I serial scores observed one year apart.[25]

Table 1 shows the results of the transition analysis between score intervals from tenure year one to two. As hypothesized, patients that rarely reported falls had the highest category stability, with nearly 84% of patients in the [.5, 1] interval falling in that same interval in year two. Though constituting a relatively small sample, patients who most often reported falls in year one had the second highest relative transition frequency to the same interval in year two, at 66.7%.

Table 3. Mean score range transitions from year one to year two. A mean score of +1.0 would be a “perfect score” of no reported falls, a mean score of -1.0 indicates that all responses indicated a recent fall had occurred, and scores in between span the interval (0, 1). Notably, the one patient that moved from [-1.0, -.5) to [.5, 1] had a score of exactly 0.5 (the boundary) with only six measurements in year two.

| Year 1 | Year 2 | Frequency | Percent (%) |
|---|---|---|---|
| | | | |
| **Frequent Fallers** | **[-1.0, -0.5)** | 12 | 66.7% |
| | **[-0.5, 0.5)** | 5 | 27.8% |
| **[-1.0, -0.5)** | **[0.5 to 1.0]** | 1 | 5.6% |
| **Occasional Fallers** | **[-1.0, -0.5)** | 5 | 10.2% |
| | **[-0.5, 0.5)** | 24 | 49.0% |
| **[-0.5, 0.5)** | **[0.5 to 1.0]** | 20 | 48.8% |
| **Rare Fallers** | **[-1.0, -0.5)** | 1 | 0.6% |
| | **[-0.5, 0.5)** | 26 | 15.6% |
| **[0.5, 1]** | **[0.5 to 1.0]** | 140 | 83.8% |

## Validity

### *Covariates*

All 234 patients with at least 5 responses to Q_FALL had both age and sex recorded in the application database. Two patients had a recorded date of birth within two years of the database added date and these datapoints were omitted. For the remaining 232 patients, 153 had an SSRI or an SNRI scheduled for dispense via spencer during the observation window. The GEE linear model results are presented in Table 4.

Table 4. GEE linear model summary for sex, age, and whether the patient was prescribed an SSRI or SNRI during the observation window.

| Coefficient | Estimate | Std. Error | Wald | *P* value |
|---|---|---|---|---|
| | | | | |
| Intercept | 0.684 | 0.127 | 28.857 | <0.001[a] |
| Sex = ‘male’ | -0.030 | 0.069 | 0.195 | 0.66[b] |
| Patient Age | 0.001 | 0.002 | 0.092 | 0.76[b] |
| SSRI / SNRI | -0.232 | 0.062 | 14.092 | <0.001[b] |

[a]The *P* value corresponding to the hypothesis of the intercept being zero is included by convention but is not a meaningful statistic.
[b]These *P* values correspond to the two-sided test of the hypothesis of a zero regression coefficient.

The demographic factors of biological sex and age were not significantly associated with Q_FALL in this chronic care population. The presence of an SSRI or SNRI medication, however, was strongly associated with Q_FALL, with the negative sign indicating that there were more reported falls in the SSRI / SNRI group adjusted for sex and age.

### *Contemporaneous Outcomes from the spencer*

The contemporaneous associations between theoretically related spencer questions are presented in Table 5. All the associations were positive and highly significant.

Table 5. Contemporaneous association between survey questions administered on spencer.

| Question | N patients[a] | Kendall’s Tau | *P* value[b] |
|---|---|---|---|

| | | | |
|---|---|---|---|
| Rate your recent quality of life | 233 | 0.15 | <0.001 |
| How is your emotional health today? | 233 | 0.21 | <0.001 |
| How would you rate your physical health today[sic] | 264 | 0.18 | <0.001 |
| Rate your ability to perform activities today | 232 | 0.23 | <0.001 |
| Are you able to accomplish what you have planned today? | 197 | 0.18 | <0.001 |
| Hospital, ER or Urgent Care in the past month? | 192 | 0.20 | <0.001 |



## Discussion

## Principal Findings

In contrast with remote digital studies which have been plagued by poor retention, the spencer platform kept over half its patients into the third year of tenure. For the 82% of patients who continued to receive surveys at the end of their first year of tenure, the aggregate response rate was 84%, with two thirds of those patients answering at near perfect rates through their entire first year of tenure. These results suggest that patients interact with in-home medication dispensing technology differently than they do with mobile applications on smartphones. We hypothesize that smartphone users process dozens or more notifications and alerts each day and it is becoming increasingly difficult to capture patients' attention during standard smartphone engagement.

Survey engagement is only a fraction of the total engagement that a patient regularly has with the

spencer device. Considering refill loading, medication dispensing, and survey completion activities, the users from the reliability study interacted with spencer 11.4 times per day on average or over 340 separate interactions in a month. We believe that these brief, repeated engagements between the patient and the spencer unit are self-reinforcing.

It is difficult to find an exact comparison for our observed response rates. In the context of primary care research studies, Booker et al described 80% or higher as "considered excellent"[30] but these surveys are multi-item and administered at a single point in time. Ecological momentary assessment (EMA) is a survey research method that addresses phenomena as they occur,[31] which has largely transitioned from paper to electronic diaries.[32] Reported compliance with paper diaries was typically high, often near 90%, though a study that was able to detect when patients opened the diaries concluded that almost 90% of the submitted diary cards were falsified for an actual compliance of only 11%. Compliance rates in pre-smartphone electronic diary studies are also often reported to be near 90%, though rates as low as 50% have been recorded. Given the low actual response rates in the paper diary studies, this either suggests substantial benefit of a dedicated electronic survey mechanism, or some other optimism in compliance reporting.[31]

Noncoverage in internet surveys, defined as lack of access to the internet, is thought to be a more serious problem than nonresponse, or an unwillingness to participate.[33] In the case of the spencer population, noncoverage can be viewed in two different ways. Since every participant has a connection to the internet through the device itself (the spencer units have both cell and WiFi connections), there is no noncoverage in the sense of lack of internet access (though machines do go offline for periods). On the other hand, the fleet itself is comprised of polychronic patients with conditions that are serious enough to be enrolled in a care management program. This presents a challenge when the goal is to make inferences to a national population, but it is ideal for inferences about a polychronic patient population under nationally supported care management.

Nonresponse consisted of either opt outs or not answering the questions despite an on-screen presentation. We suspect survey fatigue[34] accumulates, as the same question battery has been asked to some patients for over a year. Interventions to better communicate the value of the responses to the patient as well as new mechanisms to collect and disseminate survey information are being investigated. Generative AI agents offer a promising avenue for scaling a conversational dialogue about a patient's responses.

The correlation between a patient's average response to Q_FALL in two subsequent years suggests that the spencer questions can reliably measure stable patient characteristics are correlations are consistent with published literature. For constructs that are expected to change over shorter durations of time, dynamic analysis would be necessary. Le et al discussed "transient error" which arises from fluctuations in mood.[40] Ozkok et al discussed memory activation and affective priming, a cognitive and emotional mechanism, respectively, and how adding autoregressive terms to a factor analysis could account for these.[41]

Averaging scores allowed us to work on a continuum where rare fallers and never-fallers are close, a notion supported by similarities between these groups in a one-year cohort study. This was explained by high functioning older adults falling, putting themselves in high fall risk situations.[35] The measure correlated with average response values for spencer questions that, at face value, represented known risk factors of falling. Additionally, the measure correlated with the prescription of SSRIs or SNRIs, a known risk factor for falls in older populations. Together, we take this as preliminary but compelling evidence of validity the Q_FALL derived measure.

However, the demographic factors of age and biological sex unexpectedly showed no evidence of association with Q_FALL in this study. A recent meta-analysis found that 70 of 74 studies (95%) reported gender or sex differences in fall-related outcomes,[36] and increased age is a canonical risk factor for falls.[37] We hypothesize that in a polychronic population taking many medications, such demographic factors may be weaker predictors of falling than in the general population.

While smartphone-based longitudinal surveys have been planned with varying levels of success, we are not aware of completed work that investigates the psychometric properties of health-related surveys when reformatted for mobile administration. Mareva, et al published a study protocol where a smartphone application gathered data on real-time health related quality-of-life (HRQoL) using an adaptive version of the WHOQOL-BREF questionnaire. The protocol proposed to conduct a psychometric analysis, including measurement invariance through time, and attempts to establish validity, but we were unable to determine whether the study ever took place.[38] In another study, Mack et al used a mobile app to administer repeated instances of the PHQ-2 and GAD-2 scales to college students during the COVID-19 pandemic but did not investigate psychometric properties of the mobile longitudinal surveys. This study demonstrated peaks in anxiety and depression that aligned with stressful societal events but did not address reliability or validity impacts of translating the PHQ-2 or GAD-2 to the mobile phone and with repeated measures.[39]

## Limitations

This research focused on single question measures, in which one single question was asked at a specific time (i.e. the medication dispense), where responses were averaged over one- or two-year time intervals without further weighting. The use of multiple questions in a scale is standard practice in psychometric research[24] and has the potential to increase the reliability of the resulting scale. Like the challenges of addressing test-retest reliability, the concept of “internal consistency” reliability is less clear when all questions of the scale are not asked during the same survey.

Part of the validity analysis focused on correlations between questions asked on the spencer unit. Though these questions had apparent face validity with their corresponding constructs, they themselves have not been validated. It is possible that the correlations observed were partially due to similar positioning of the responses on the device from most to least positive sentiment.

Finally, this study did not consider sensitivity to change, which is important in the context of RWE because it allows researchers and clinicians to detect change resulting from a minimal intervention.[42]

## Conclusions

This research has shown that longitudinal surveys administered via spencer, a medication dispensing smart hub, is an effective way to generate meaningful real-world data from patients in their natural environments. Patients are compliant on the spencer platform at a much higher rate that mobile apps and persist on platform for years. The measures derived from the longitudinal surveys can be reliable and valid measures of important health constructs.

Because medication dispensing is a fundamental component of the survey generating mechanism, real-world data from such a platform offers an ideal opportunity to study medication effectiveness or health outcomes. It is a tool for generating real world evidence to support new indications for already

approved drugs or to demonstrate relationships between health outcomes and economic factors.

## Acknowledgements

This work was funded by Spencer Health Solutions, Inc., a private corporation headquartered in Morrisville, North Carolina. The authors wish to thank University of Maryland Professor Greg Hancock for his perspective on presenting this new source of patient reported outcomes. We thank Devi Kodavanti, a member of the Data team at Spencer Health Solutions, Inc., for computing many reliability coefficients in the past that gave us confidence to do this research. Our resident designer Lew Wilson created the TOC image. Finally, we thank Alan Menius, the scientist behind the evolution of the spencer standard questions.

## Conflicts of Interest

The authors are all current or former employees of Spencer Health Solutions, Inc., the developer of the spencer platform.

## Abbreviations

EMA: European Medicines Agency, Ecological Momentary Assessment
FDA: US Food and Drug Administration
FES-I: Falls Efficacy Scale-International
GEE: Generalized Estimating Equation
JMIR: Journal of Medical Internet Research
PRO: Patient Reported Outcome
PROMIS: Patient-Reported Outcomes Measurement Information System
Q_FALL: question about recent falls asked on spencer
RR: Response Rate
RWD: Real-World Data
RWE: Real-World Evidence
SSRI: Selective Serotonin Reuptake Inhibitor
SNRI: Serotonin and Norepinephrine Reuptake Inhibitors

## References

1. Huang YL, Moon J, Segal JB. A Comparison of Active Adverse Event Surveillance Systems Worldwide. *Drug Saf*. 2014;37(8):581-596. doi:10.1007/s40264-014-0194-3
2. FDA. *Framework for FDA's Real-World Evidence Program*.; 2018. Accessed March 13, 2024. https://www.fda.gov/media/120060/download
3. Klonoff DC. The New FDA Real-World Evidence Program to Support Development of Drugs and Biologics. *J Diabetes Sci Technol*. 2020;14(2):345-349. doi:10.1177/1932296819832661
4. *Considerations for the Use of Real-World Data and Real-World Evidence To Support Regulatory Decision-Making for Drug and Biological Products*.; 2023.
5. EMA. *Real-World Evidence Framework to Support EU Regulatory Decision-Making: Report on the Experience Gained with Regulator-Led Studies from September 2021 to February 2023*.; 2023. Accessed March 13, 2024. https://www.ema.europa.eu/system/files/documents/report/real-world-evidence-framework-support-eu-regulatory-decision-making-report-experience-gained_en.pdf
6. CADTH. *Guidance for Reporting Real-World Evidence*.; 2023. Accessed March 13, 2024.

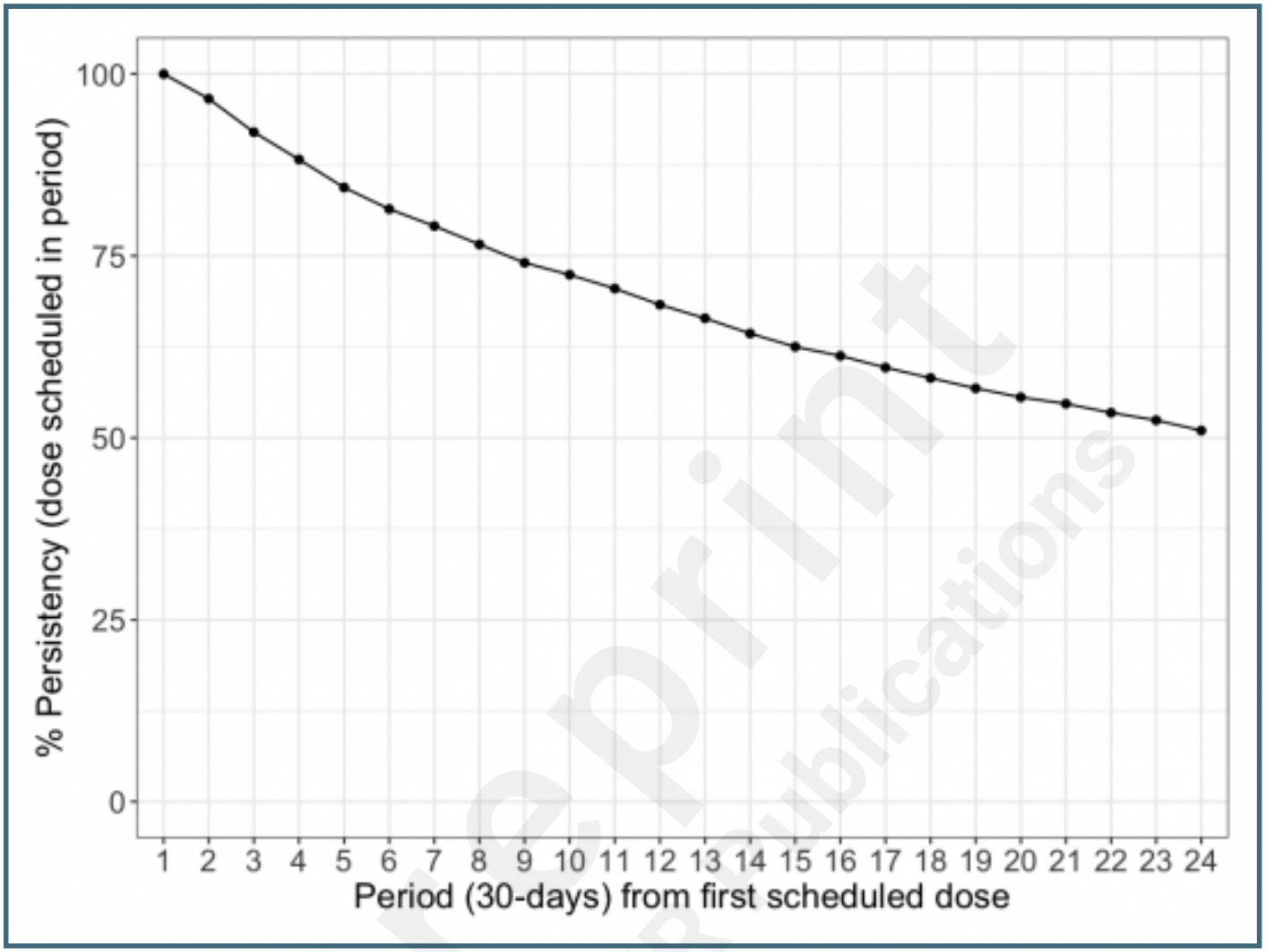
https://www.cadth.ca/sites/default/files/RWE/MG0020/MG0020-RWE-Guidance-Report-Secured.pdf
7. Purpura CA, Garry EM, Honig N, Case A, Rassen JA. The Role of Real-World Evidence in FDA-Approved New Drug and Biologics License Applications. *Clin Pharmacol Ther*. 2022;111(1):135-144. doi:10.1002/CPT.2474
8. *Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*.; 2009.
9. Hufford MR, Shiffman S. Assessment Methods for Patient-Reported Outcomes. *Disease Management & Health Outcomes*. 2003;11(2):77-86. doi:10.2165/00115677-200311020-00002
10. Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. *NPJ Digit Med*. 2018;1(1):18. doi:10.1038/s41746-018-0029-1
11. Hillygus DS, Snell S. *Longitudinal Surveys: Issues and Opportunities*. Vol 1. (Atkeson LR, Alvarez RM, eds.). Oxford University Press; 2015. doi:10.1093/oxfordhb/9780190213299.013.7
12. Lynn P. *Methodology of Longitudinal Surveys*.; 2009.
13. Göritz AS, Reinhold N, Batinic B. Online panels. *Online social sciences*. Published online 2002:27-47.
14. Hays RD, Liu H, Kapteyn A. Use of Internet panels to conduct surveys. *Behav Res Methods*. 2015;47(3):685-690. doi:10.3758/s13428-015-0617-9
15. Lin YK, Newman S, Piette J. Response Consistency of Crowdsourced Web-Based Surveys on Type 1 Diabetes. *J Med Internet Res*. 2023;25:e43593. doi:10.2196/43593
16. Pratap A, Neto EC, Snyder P, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. *NPJ Digit Med*. 2020;3(1):21. doi:10.1038/s41746-020-0224-8
17. Senoo K, Miki T, Ohkura T, et al. A Smartphone App to Improve Oral Anticoagulation Adherence in Patients With Atrial Fibrillation: Prospective Observational Study. *JMIR Mhealth Uhealth*. 2022;10(1):e30807. doi:10.2196/30807
18. Patel T, Ivo J, Pitre T, Faisal S, Antunes K, Oda K. An In-Home Medication Dispensing System to Support Medication Adherence for Patients With Chronic Conditions in the Community Setting: Prospective Observational Pilot Study. *JMIR Form Res*. 2022;6(5):e34906. doi:10.2196/34906
19. McBride CM, Morrissey EC, Molloy GJ. Patients' Experiences of Using Smartphone Apps to Support Self-Management and Improve Medication Adherence in Hypertension: Qualitative Study. *JMIR Mhealth Uhealth*. 2020;8(10):e17470. doi:10.2196/17470
20. Palmer MJ, Machiyama K, Woodd S, et al. Mobile phone-based interventions for improving adherence to medication prescribed for the primary prevention of cardiovascular disease in adults. *Cochrane Database of Systematic Reviews*. 2021;2021(3). doi:10.1002/14651858.CD012675.pub3
21. Allison PD. Discrete-Time Methods for the Analysis of Event Histories. *Sociol Methodol*. 1982;13:61. doi:10.2307/270718
22. Yardley L, Beyer N, Hauer K, Kempen G, Piot-Ziegler C, Todd C. Development and initial validation of the Falls Efficacy Scale-International (FES-I). *Age Ageing*. 2005;34(6):614-619. doi:10.1093/ageing/afi196
23. Chen Y, Du H, Song M, et al. Relationship between fear of falling and fall risk among older patients with stroke: a structural equation modeling. *BMC Geriatr*. 2023;23(1):647. doi:10.1186/s12877-023-04298-y
24. JHANGIANI RS, CHIANG ICA, CUTTLER C, LEIGHTON DC. *Research Methods in Psychology*. 4th edition. Kwantlen Polytechnic University; 2019.
25. Delbaere K, Close JCT, Mikolaizak AS, Sachdev PS, Brodaty H, Lord SR. The Falls Efficacy Scale International (FES-I). A comprehensive longitudinal validation study. *Age Ageing*. 2010;39(2):210-216. doi:10.1093/ageing/afp225
26. Dewan N, MacDermid JC. Fall Efficacy Scale - International (FES-I). *J Physiother*. 2014;60(1):60. doi:10.1016/j.jphys.2013.12.014

27. Darowski A, Chambers SACF, Chambers DJ. Antidepressants and Falls in the Elderly. *Drugs Aging*. 2009;26(5):381-394. doi:10.2165/00002512-200926050-00002
28. Park H, Satoh H, Miki A, Urushihara H, Sawada Y. Medications associated with falls in older people: systematic review of publications from a recent 5-year period. *Eur J Clin Pharmacol*. 2015;71(12):1429-1440. doi:10.1007/s00228-015-1955-3
29. Halekoh U, Højsgaard S, Yan J. The R Package geepack for Generalized Estimating Equations. *J Stat Softw*. 2006;15(2). doi:10.18637/jss.v015.i02
30. Booker QS, Austin JD, Balasubramanian BA. Survey strategies to increase participant response rates in primary care research studies. *Fam Pract*. 2021;38(5):699-702. doi:10.1093/fampra/cmab070
31. Shiffman S, Stone AA, Hufford MR. Ecological Momentary Assessment. *Annu Rev Clin Psychol*. 2008;4(1):1-32. doi:10.1146/annurev.clinpsy.3.022806.091415
32. Csikszentmihalyi M, LeFevre J. Optimal experience in work and leisure. *J Pers Soc Psychol*. 1989;56(5):815-822. doi:10.1037/0022-3514.56.5.815
33. Couper MP, Kapteyn A, Schonlau M, Winter J. Noncoverage and nonresponse in an Internet survey. *Soc Sci Res*. 2007;36(1):131-148. doi:10.1016/j.ssresearch.2005.10.002
34. de Koning R, Egiz A, Kotecha J, et al. Survey Fatigue During the COVID-19 Pandemic: An Analysis of Neurosurgery Survey Response Rates. *Front Surg*. 2021;8. doi:10.3389/fsurg.2021.690680
35. Asai T, Oshima K, Fukumoto Y, Yonezawa Y, Matsuo A, Misu S. The association between fear of falling and occurrence of falls: a one-year cohort study. *BMC Geriatr*. 2022;22(1):393. doi:10.1186/s12877-022-03018-2
36. Sebastiani C, Wong JYX, Litt A, et al. Mapping sex and gender differences in falls among older adults: A scoping review. *J Am Geriatr Soc*. 2024;72(3):903-915. doi:10.1111/jgs.18730
37. Ambrose AF, Paul G, Hausdorff JM. Risk factors for falls among older adults: A review of the literature. *Maturitas*. 2013;75(1):51-61. doi:10.1016/j.maturitas.2013.02.009
38. Mareva S, Thomson D, Marenco P, et al. Study Protocol on Ecological Momentary Assessment of Health-Related Quality of Life Using a Smartphone Application. *Front Psychol*. 2016;7. doi:10.3389/fpsyg.2016.01086
39. Mack DL, DaSilva AW, Rogers C, et al. Mental Health and Behavior of College Students During the COVID-19 Pandemic: Longitudinal Mobile Smartphone and Ecological Momentary Assessment Study, Part II. *J Med Internet Res*. 2021;23(6):e28892. doi:10.2196/28892
40. Le H, Schmidt FL, Putka DJ. The Multifaceted Nature of Measurement Artifacts and Its Implications for Estimating Construct-Level Relationships. *Organ Res Methods*. 2009;12(1):165-200. doi:10.1177/1094428107302900
41. Ozkok O, Zyphur MJ, Barsky AP, Theilacker M, Donnellan MB, Oswald FL. Modeling Measurement as a Sequential Process: Autoregressive Confirmatory Factor Analysis (AR-CFA). *Front Psychol*. 2019;10. doi:10.3389/fpsyg.2019.02108
42. Hays RD, Hadorn D. Responsiveness to change: an aspect of validity, not a separate dimension. *Quality of Life Research*. 1992;1(1):73-75. doi:10.1007/BF00435438

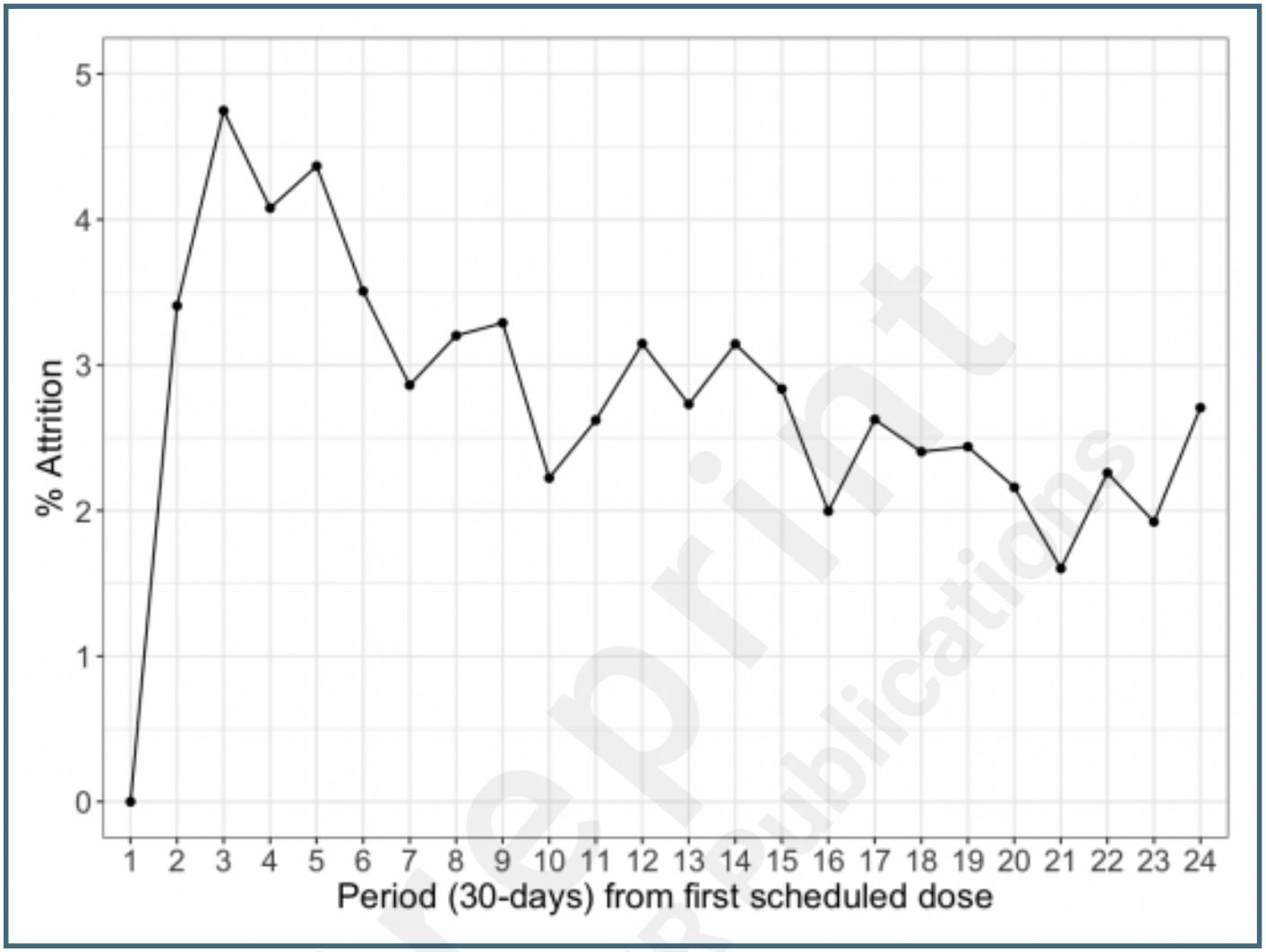
# Supplementary Files

# Figures

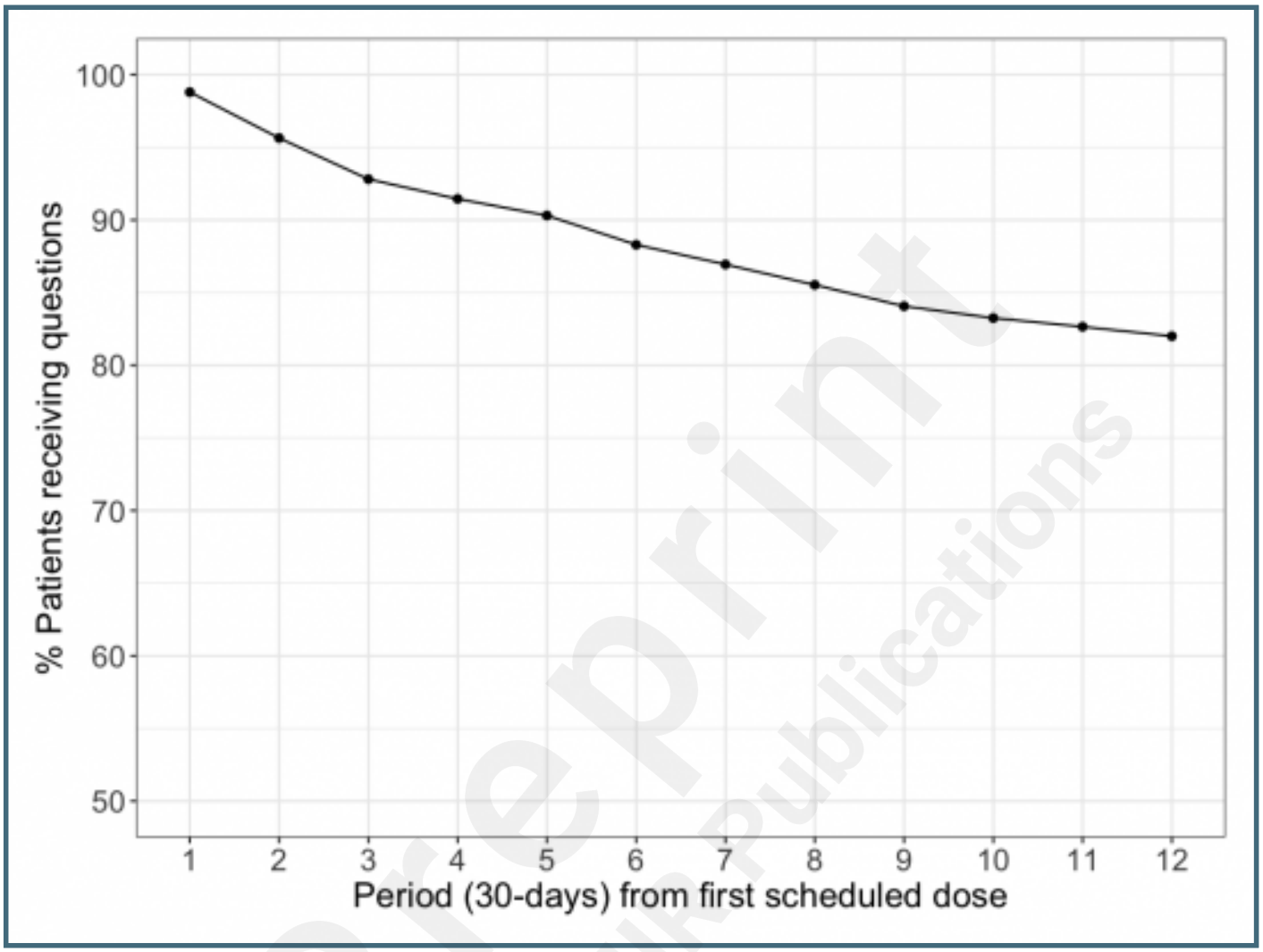
Platform persistency over the first two years of patient tenure.



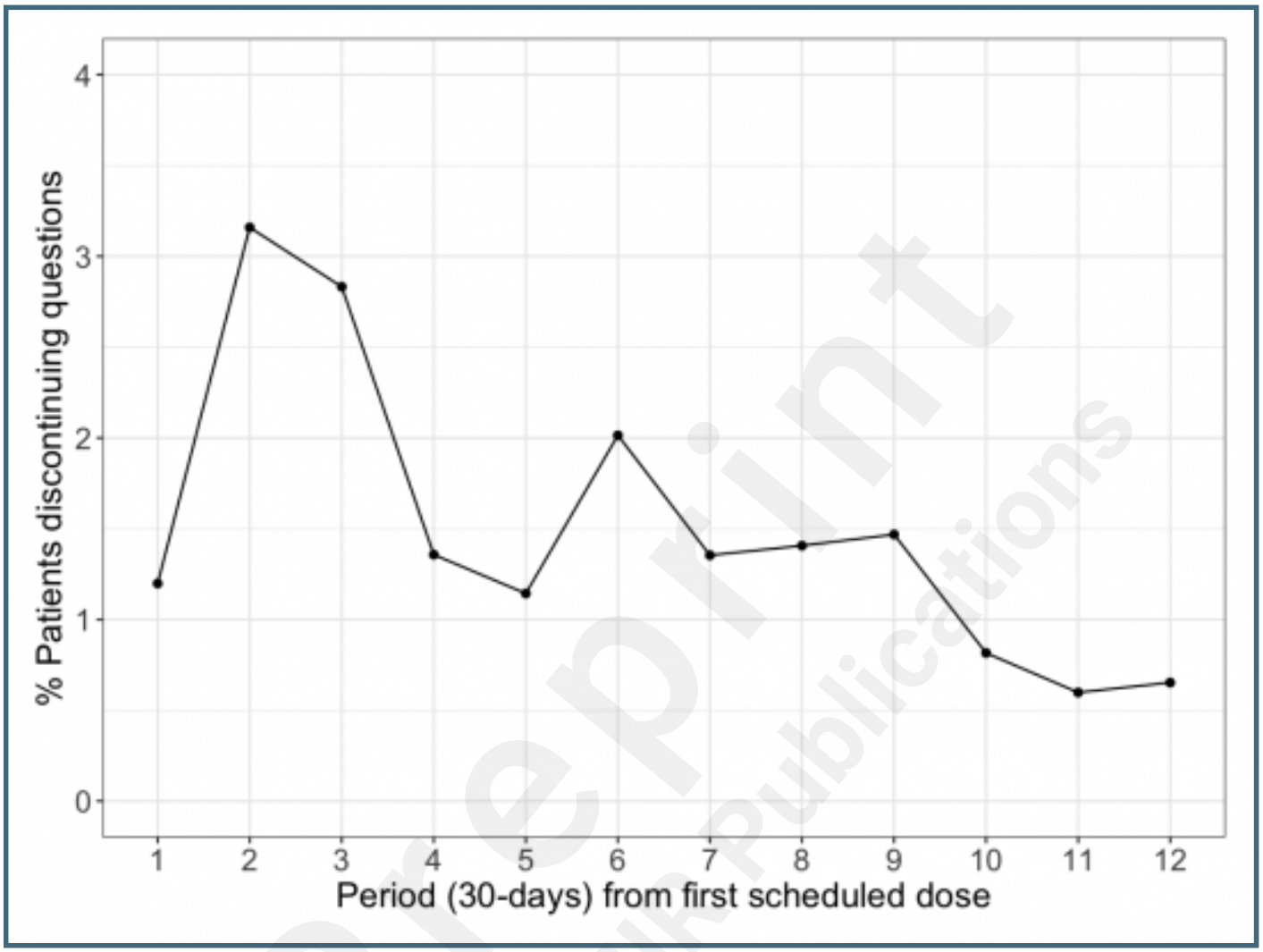
Platform attrition rate over the first two years of patient tenure.



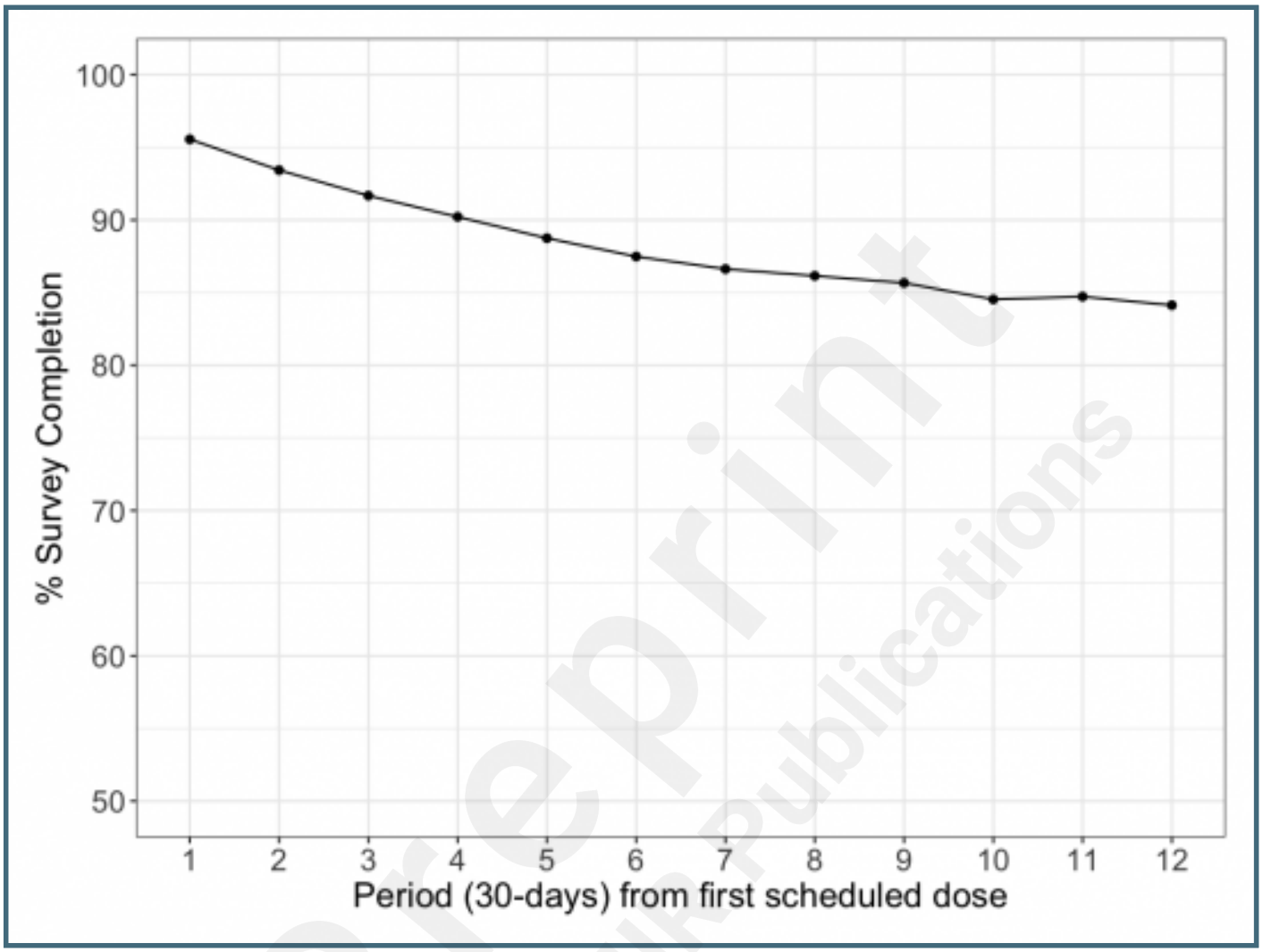
Proportion of patients who continue to receive surveys during their first year of tenure on the spencer platform.



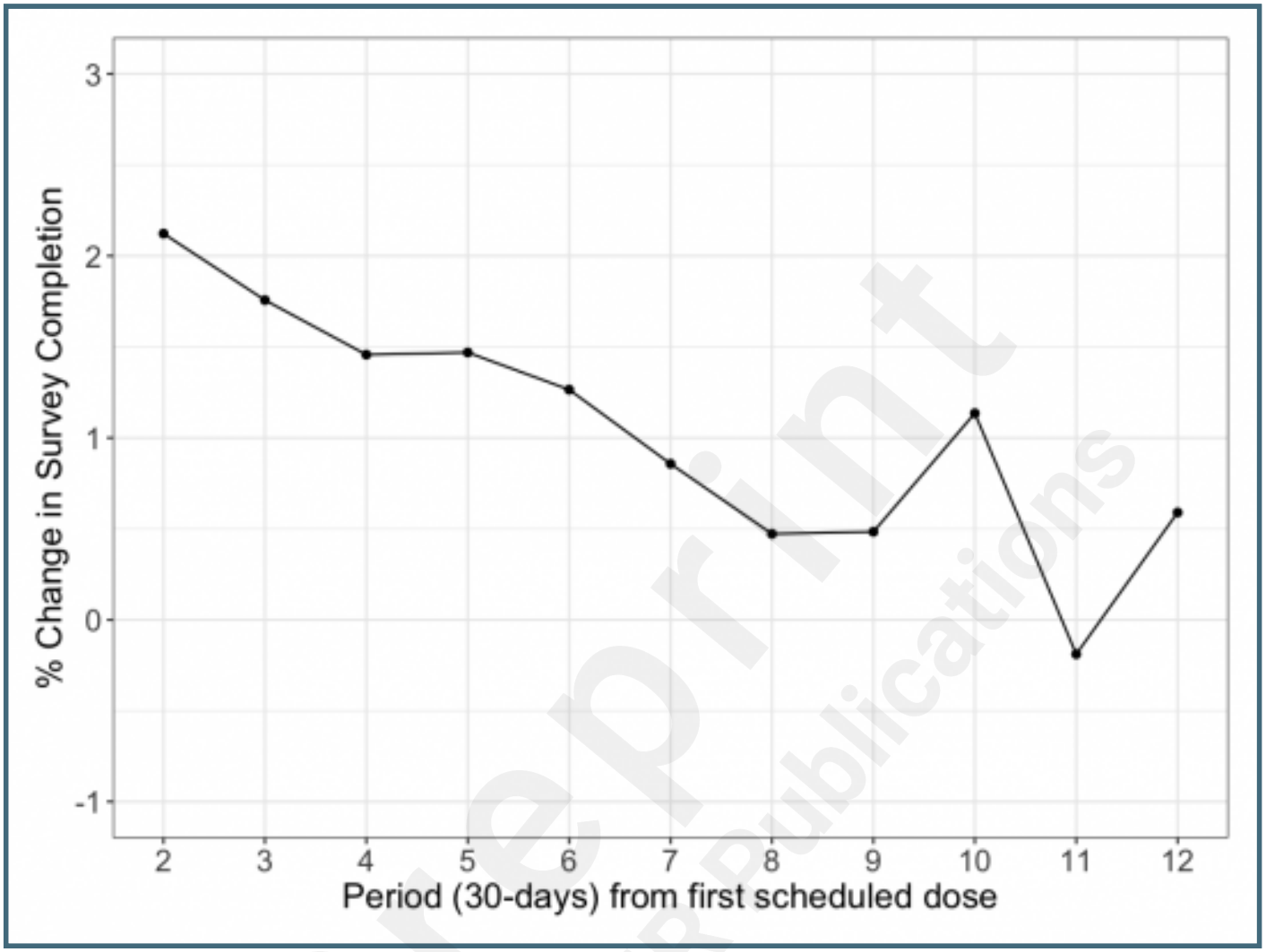
Discontinuation rate for survey questions during the patient’s first year of tenure on the spencer platform.



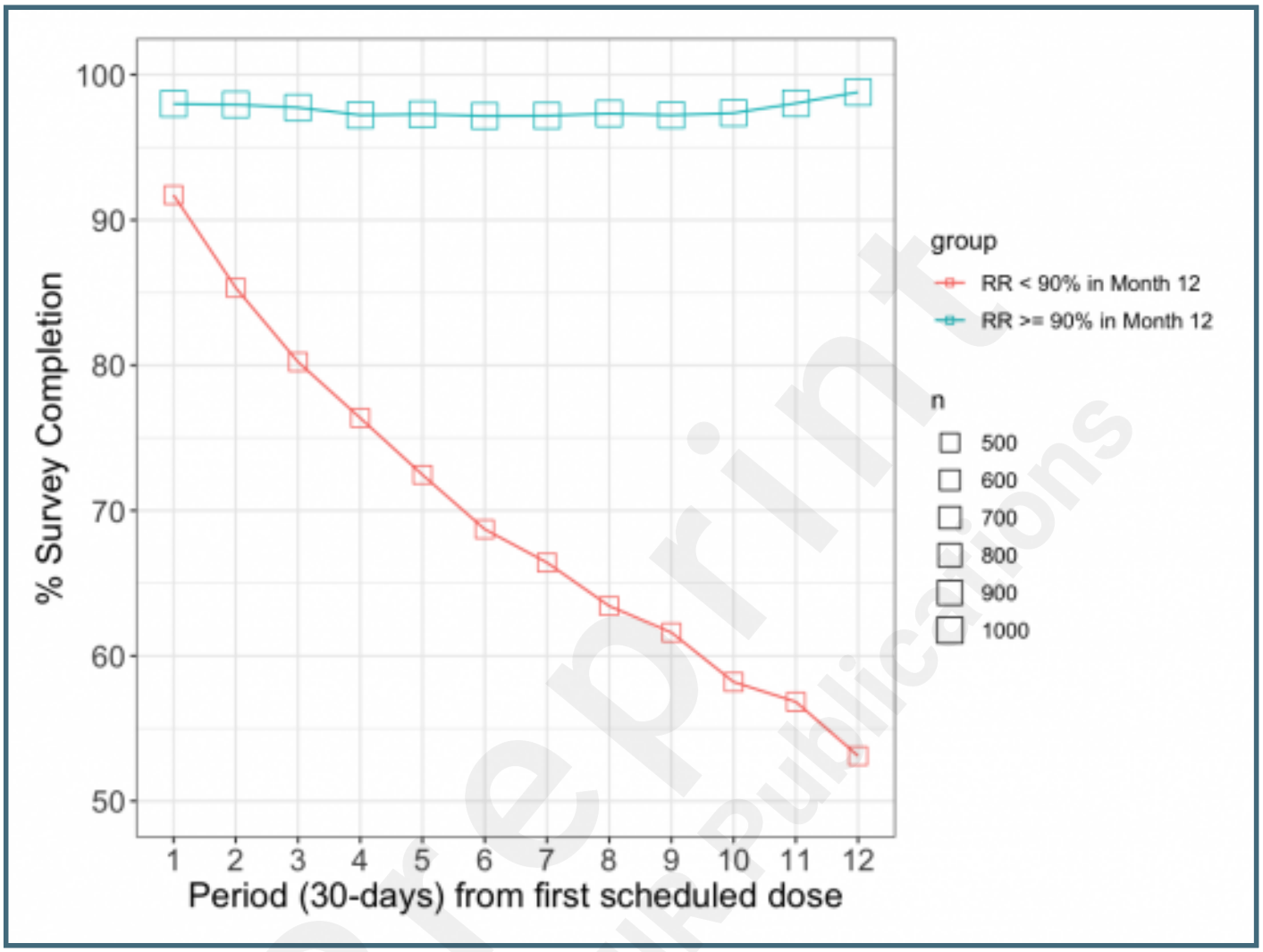
Proportion of surveys that are responded to during the first year of tenure among patients who stay on platform.



Change in survey completion rate over the first year of tenure.



Proportion of surveys that are completed, grouped by whether the survey response rate (RR) was >=90% or < 90% in period 12.



# TOC/Feature image for homepages

Image of a patient holding a dispensed medication pouch and looking at a displayed question on the spencer unit.