

# Home-based Digital Assessments with Applied Sentiment & Emotion AI Capture Improved Quality-of-life in Asthma Patients

Bo Zhang<sup>1</sup>, Ruben Buendia<sup>1</sup>, Nicholas Iannotti<sup>1</sup>, Emma Ramsden<sup>2</sup>, Paul O'Regan<sup>3</sup>, Jason Swift<sup>4</sup>,  
Sarah Lockwood<sup>5</sup>, David J. Jackson<sup>6</sup>, Glynn Dennis<sup>1</sup>, Lynn Hagger<sup>7</sup> and Jesper Havsol<sup>1</sup>

**Abstract**—With the rise of digital transformation in the pharmaceutical industry, digital therapeutics are being integrated in drug development clinical trials. In the TWINKLE study, information about asthmatic patients' disease control and quality-of-life (QoL) was measured by daily video recording, in conjunction with daily electronic questionnaires and home-based spirometry. From the video messages, sentiment and emotion AI was applied to detect subtle QoL changes in asthmatic patients after receiving treatments. Sentiment scores, derived from patients' daily messages via natural language processing, correlated strongly with metrics of lung functions and outcomes of electronic questionnaires. However, video-derived emotional analysis exhibited strong interpersonal variations and systematic biases, yet still showed utility in detecting QoL changes after personalized calibration and signal aggregation. Compared to traditional patient-reported outcomes, all three categories of digital measurements were able to detect significantly improved asthma control from patients who responded to treatments. The result provides insights into developing novel digital outcomes through the application of connected digital devices and advanced AI tailored to clinical settings.

**Clinical relevance**— Digital outcomes involving connected digital devices and AI for sentiment/emotion analysis could capture subtle QoL changes reliably and earlier than hospital visits, reducing burden and improving disease management. Integrating digital therapeutics in asthma drug development trials may prove to be feasible and valuable.

## I. INTRODUCTION

AstraZeneca is living a Digital Transformation in which digital therapeutics play an important role. This transformation has reached respiratory therapeutic development, and more specifically, asthma drug development. In this context, the TWINKLE study (NCT04200326) explored novel technologies to collect information about the patient experience before and early after starting benralizumab (Fasenra) as standard of care (SoC) for severe uncontrolled asthma in a real-world setting, to determine how the experience changes over time.

The primary objective was changes in quality of life (QoL) from baseline to 4 weeks after first dose of benralizumab,

\*This work was supported by AstraZeneca. <sup>1</sup>B.Z., R.B., N.I., G.D., and J.H. are with Data Science & Artificial Intelligence, BioPharmaceuticals RD, AstraZeneca. <sup>2</sup>E.R., is with Evidence Delivery & Operations, Global Medical Affairs, BioPharmaceuticals Business, AstraZeneca. <sup>3</sup>P.O., • digital Experimental Cancer Medicine Team, CRUK Manchester Institute Cancer Biomarker Centre, Manchester, UK. <sup>4</sup>J.S., is with Precision Medicine, Oncology RD, AstraZeneca. <sup>5</sup>S.L., is with late R&I BioPharmaceuticals R&D, AstraZeneca. <sup>6</sup>D.J.J. is with the School of Immunology & Microbial Sciences, King's College London, and with the Guy's and St Thomas' NHS Foundation Trust. <sup>7</sup>L.H. is with Content Strategy & Experience Design, Digital Global Commercial, Business, AstraZeneca. Corresponding author J.H. (jesper.havsol@astrazeneca.com)

as captured by novel digital assessments. The study was designed to capture data very early on because patients have informally reported feeling physically and mentally better a few days after the first dose of benralizumab. Such feelings were highly individualized to each patient and were rarely captured in a clinical setting. Here we investigate whether home-based digital measures can capture the signs of QoL changes earlier than the hospital visits in four weeks. Changes in QoL were measured by using video capture and facial analysis technology in conjunction with a daily electronic QoL questionnaire and spirometry.

## II. BACKGROUND

Phase III trials for asthma drugs are often very large, expensive, and burdensome for patients. For instance, the two Phase III trials evaluating benralizumab together recruited over 5 000 patients and lasted for almost a year. Nowadays, connected devices, mobile sensors, smart phone apps, and AI algorithms among other novel technologies enable digital data capture and assessments, promising more informative trial outcomes based on the high frequency and density of the data provided by such out-of-clinic monitoring. These would enable trials which could be less burdensome, shorter, and require fewer patients [1]. TWINKLE is a step forward in this direction, but being a pilot study, it involved few patients. Therefore, detection of an early first-dose effect will be further validated in a larger follow-up study.

## III. METHODS

### A. The TWINKLE study

The endpoints of TWINKLE included: improvement in QoL measures, physical and mental well-being as assessed by an increase in positive facial expressions, positive keywords, and quality of life biomarkers, and improvement in peak expiratory flow (PEF) and forced expiratory volume in one second (FEV1) within the first 4 weeks after beginning benralizumab. Approximately 10 patients aged 18-75 years with severe uncontrolled asthma beginning benralizumab as SoC were to be recruited from one site in London, England. Unfortunately, only 5 patients were recruited before early closing due to COVID-19, and 1 of those patients was not eligible.

Each patient visited the study site 3 times: enrolment, benralizumab injection, and last visit, and performed daily tasks at home. Electronic medical records (EMR) were taken at the first visit. Daily data from 3 categories of devices (electronic questionnaire, spirometry, and video messaging)

were recorded for 2 weeks prior and 4 weeks after treatment (benralizumab injection). Information about traditional PROs was registered at hospital visits. The procedure is shown in Fig. 1, and the digital devices deployed in Fig. 2.

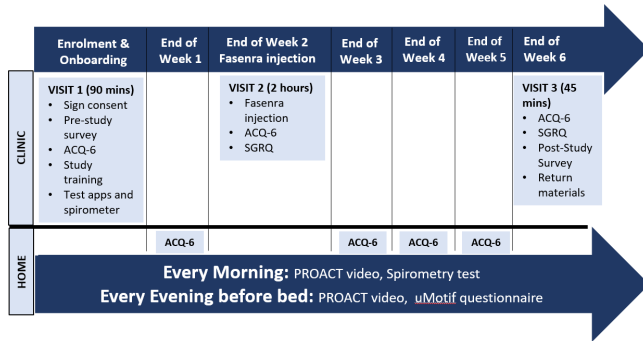


Fig. 1. Study procedure

### B. Traditional Patient-Reported Outcomes (PRO)

The Asthma Control Questionnaire 6 (ACQ-6) assesses asthma symptoms (night-time waking, symptoms on waking, activity limitation, shortness of breath, wheezing, and rescue medication use) in a period of one week [2]. The ACQ scores have a range from 0 (no symptoms) to 6 (maximum impairment). A change of 0.5 is considered clinically significant. Analogously, St. George's Respiratory Questionnaire (SGRQ) is a 50-item PRO to estimate health status of patients with airway obstruction diseases [3]. SGRQ scores can range from 0 (no symptoms) to 100 (maximum impairment), and a change of 4 is considered clinically significant.

### C. uMotif Electronic QoL Questionnaire

uMotif (uMotif, London, UK) is a configurable health data capture platform. uMotif mobile application posed questions about daily disease control and general feelings: inhaler, breathless, wheeze, cough, wake up last night, exercise, energy, stress, activities, mood, impacts on QoL, and motivation. Each question score range from 1 (best) to 5 (worst). The average score of 10 user-selected questions was used as daily ePRO metric.

### D. Home-based Smart Spirometry

Patients' lung air capacity, specifically, PEF, FEV1, and forced vital capacity (FVC) were measured daily through MIR Spirobank Smart® spirometer, additional data points such as FEV1/FVC ratios were subsequently calculated. Patients were asked to complete several repeats before filing their uMotif, and medians of daily reading were used.

### E. PROACT Video Messaging

Patients were asked to record a few seconds of video every morning and evening using PROACT [4]. These videos were then manually transcribed by a third-party. Sentiment detection on the text transcripts was performed using Microsoft Azure Media Text Analytics. Sentiment scores from 0 (most negative) to 1 (most positive) were derived from

features including n-grams, features generated from part-of-speech tags, and word embeddings. A sentiment score larger than 0.5 was deemed positive; otherwise, negative. Emotion recognition was performed using Microsoft Azure Media Analytics Face, which provided scores covering eight emotions (anger, contempt, disgust, fear, happiness, neutral, sadness and surprise) for each face in every frame of the recording. Emotional data from the eight components was scaled to the sum of one per frame and re-sampled to the second. To aggregate the eight emotional components and mitigate the biases of personal traits in emotion analysis, a predicted probability of a positive sentiment was given for each message by fitting a logistic regression model, which associated the eight emotional components with the result of sentiment analysis for each patient messages.

### F. Data Processing

Data from all sources was aligned to measurements timestamps, re-sampled to 12 hours, and aggregated taking the maximum of the corresponding time periods. For each measurement, all data recorded before benralizumab injection (day-0) were considered baseline (BL) readings. Data collected after the treatment was grouped by week and compared to BL. For visualization purposes, independent two-tail t-tests (Fisher's exact tests for sentiments) were performed comparing measurements before and after benralizumab injections. Raw p-values are reported in the figures without adjustment for multiple-testing.

## IV. RESULTS

### A. Traditional PRO Outcomes

As shown in Fig. 3, three patients' ACQ-6 decreased  $>0.5$  from BL. The same patients SGRQ score decreased  $>4$ . Given the scale of these changes, the three patients meet the criteria to be classified as responders to treatment. One of the responders improved the ACQ score before benralizumab injection due to prescription of oral corticosteroids.

### B. Digital Outcomes

Violin plots representing weekly and BL distributions of values for each digital outcome are shown in Fig. 4. All three responder patients showed improvement (decreased scores) in uMotif ePROs. Significant improvements (paired T-Test  $p < 0.05$ ) were observed the first week after benralizumab injection. Moreover, all three responders showed improvement in at least two of the three measures of lung function. PEF was the only measure that improved in all three responders by the end of study.

Regarding sentiment analysis, scores from daily PROACT video messages showed a bi-modal distribution. Most positive messages were scored close to 1, while negative messages were scored close to zero. The three responder patients provided messages scored more positively after receiving treatment than before. Significant improvements (Fisher's exact  $p < 0.05$ ) were observed 2 weeks after the treatment, similar to spirometry results.

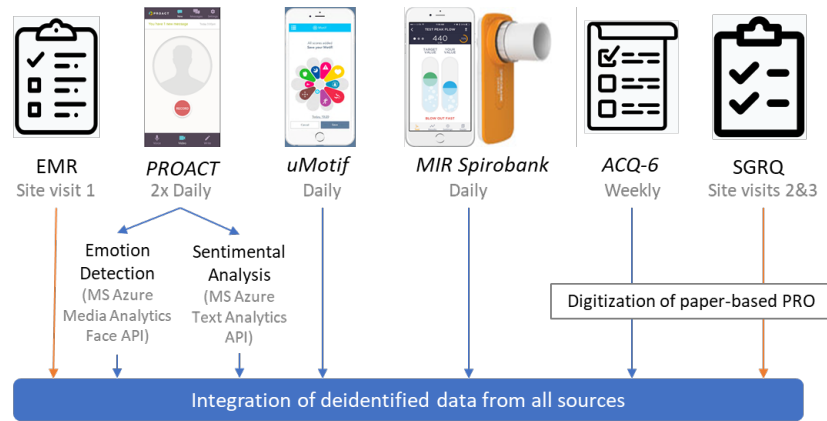


Fig. 2. Devices and instruments used in TWINKLE study

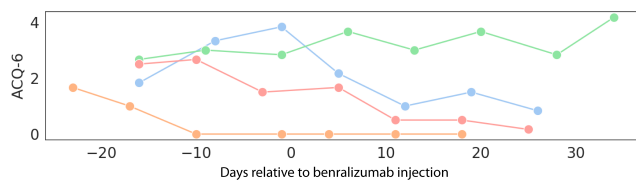


Fig. 3. Patients longitudinal ACQ-6 score, each color represents a different patient.

None of the eight emotional components captured the response to treatment. Furthermore, the association between individual emotional components in the videos and the sentiment scores of corresponding messages was weak, possibly indicating strong personal traits. Unlike sentiment scores, the emotions detected in the videos appeared to be affected by the time of the day when the videos were recorded.

The non-responding patient did not exhibit improvement in any of the digital outcomes. Further, the patient’s struggle with asthma control was reflected in the video recordings, in which only 7 messages (10%) were scored as being positive.

### C. Early Signs of Response, the “Twinkle”

To measure early signs of QoL improvement, we defined the moment of the “Twinkle” by combining data from spirometry, ePRO, and video recordings. For each of the responder patients, as early as 2 to 4 days after treatment, at least one measure of the three technologies improved compared to the upper-quartiles from baseline data. An example of a responder patient is shown in Fig. 5.

### D. Correlation Between Outcomes

Fig. 6 shows the 7-day average of uMotif and sentiment scores alongside ACQ-6 scores. Strong correlations indicate consistency in measuring QoL changes by both instruments. Spirometry data also generally correlated well with ACQ-6 at an individual level, but exhibited interpersonal variances, e.g., the FEV1/FVC ratio was not an indicator of QoL improvement for one of the responder patients.

Leveraging the granularity of daily data, we investigated the association between ePRO and sentiment results. Sur-

prisingly, we found a strong though non-linear correlation, see Fig. 7. This provides a strong indication that increased asthma severity may have a causal effect, and result in an increased number of negative messages (scores below 0.5) being recorded. Therefore, negative messages could be more informative than positive ones for measuring disease control.

## V. DISCUSSION

The TWINKLE study aimed to detect signs of improved QoL using digital technologies. Information about patients’ experience was collected daily using three technologies: spirometry, uMotif ePRO, and PROACT video capture. Three of the four patients who completed the study met the criteria to be considered responders to treatment. All three digital outcomes can detect improved asthma control as they were well correlated to ACQ-6 scores.

Daily uMotif was the only digital outcome that captured significant changes from the first week after treatment, and PEF was the spirometry measure that most significantly captured improvements. Notably, signs of improvement could be observed as early as two days after benralizumab injection.

From the PROACT daily video recording transcriptions, sentiment scores based on natural language processing captured significant QoL changes, which were not reflected in individual emotional components detected by facial recognition. Negative sentiments correlated strongly with increased symptoms as described in PROs. However, discrepancies between sentiment scores and the actual messages were found.

Data generated from emotional state analysis contained a high level of noise, as each patient seemed to have a personalized baseline of emotions. Therefore, individual-level calibration and signal aggregation are required to detect changes in QoL. The regression models developed using emotions as features to predict sentiment scores indicate that emotion scores are meaningful and do reflect the sentiment of the message being delivered. Further work is needed in exploring the personalization of emotion algorithms to the enable prediction of a patient’s quality of life.

In relation to the timing of benralizumab treatment effect onset, generalizability of the findings is limited due to the

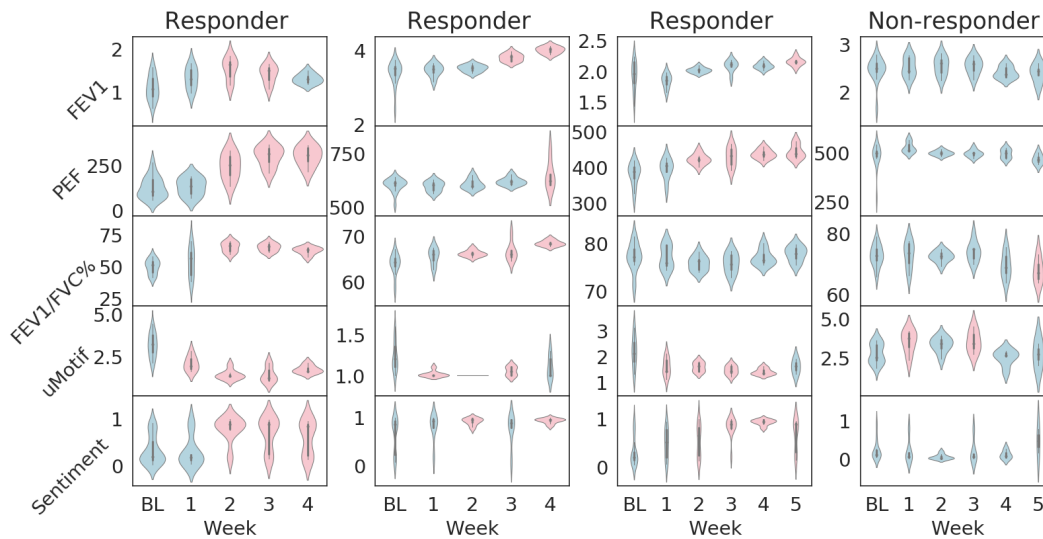


Fig. 4. Weekly summary of digital measurements, each plot represent a separate patient, 3 responders and 1 non-responder. Compared to baseline readings (BL), red “violins” indicate significant changes (raw  $p < 0.05$ )

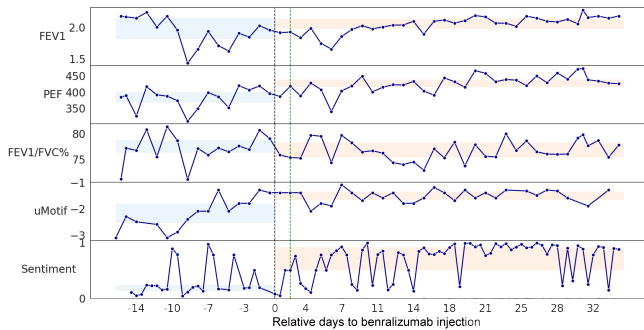


Fig. 5. Visualization of one responder patient's daily data. Higher scores represent improvements for all parameters. The earliest sign of improvement “Twinkle” is marked at day-2, where PEF, uMotif and sentiment score are improved from the upper-quartiles of baseline readings.

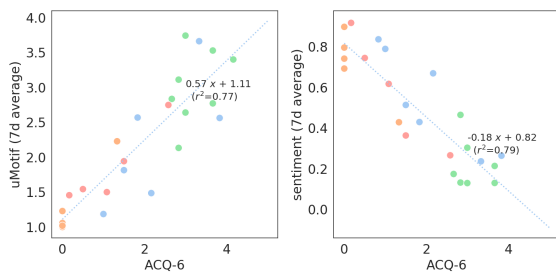


Fig. 6. Correlation between weekly summarized digital outcomes and traditional ACQ-6 PRO, each color represents a different patient.

low number of subjects. However, the results do indicate that home-based monitoring can support the early identification of treatment response.

The TWINKLE clinical study described in this paper were approved by the Queen Square Research Ethics Committee, London (REC Reference: 19/LO/1476).

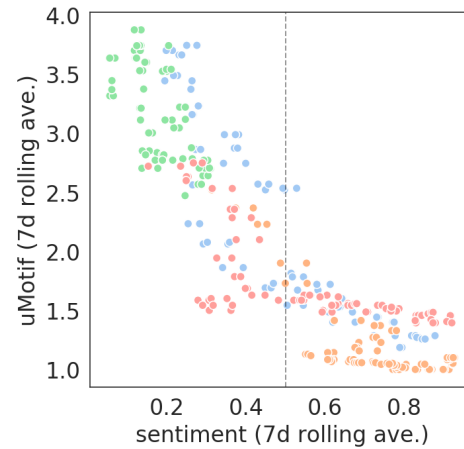


Fig. 7. Correlation between sentiment scores and ePRO, each color represents a different patient.

## ACKNOWLEDGMENT

We would like to thank the researchers at digital Experimental Cancer Medicine Team ([www.digitalecmt.org](http://www.digitalecmt.org)) for their support of the PROACT platform.

## REFERENCES

- [1] Warraich, H.J., Califf, R.M. and Krumholz, H.M., 2018. The digital transformation of medicine can revitalize the patient-clinician relationship. NPJ digital medicine, 1(1), pp.1-3.
- [2] Juniper, E.F., O'byrne, P.M., Guyatt, G.H., Ferrie, P.J. and King, D.R., 1999. Development and validation of a questionnaire to measure asthma control. European respiratory journal, 14(4), pp.902-907.
- [3] Jones, P.W., Spencer, S. and Adie, S., 2009. St George's respiratory questionnaire manual. St Georges University of London.
- [4] Hughes, A., Landers, D., Arkenau, H.T., Shah, S., Stephens, R., Mahal, A., Simmons, M., Lemech, C. and Royle, J., 2016. Development and evaluation of a new technological way of engaging patients and enhancing understanding of drug tolerability in early clinical development: PROACT. Advances in therapy, 33(6), pp.1012-1024.