Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics and Data Science

Master's thesis

Individual Reference Intervals for multiple sclerosis biomarkers

Anh Phu'o'ng Dô

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Biostatistics

SUPERVISOR :
Prof. dr. Olivier THAS
Prof. dr. ir. Liesbet PEETERS

MENTOR :
dr. Murih PUSPARUM

2023
2024

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics and Data Science

Master's thesis

Individual Reference Intervals for multiple sclerosis biomarkers

Anh Phu'o'ng Dô

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Biostatistics

SUPERVISOR :

Prof. dr. Olivier THAS

Prof. dr. ir. Liesbet PEETERS

MENTOR :

dr. Murih PUSPARUM

Acknowledgements

First and foremost, I owe an immense debt of gratitude to my advisors: Prof. Thas, Prof. Peeters, Dr. Murih, and Mr. Khan. Their constant dedication and guidance play an important role in bringing this research to fruition. Without their expertise and kindness, completing this thesis would have been an impossible challenge.

It's difficult to find the right words to express my thanks to Dr. Murih for her continuous support and counsel not only during this research but also during challenging time.

I am grateful for the support of my friends, Ilaria and Roel, who have stood by me through all the challenging moments, always believing in me and the value of my work.

I would also like to thank my husband, Brice, and my kids, Django and Yuri, who always try their best to support me in whatever I decide to do. Without their unconditional love and not-funny jokes, no study can be done.

Lastly, I extend my heartfelt thanks to Prof. Molenberghs and Prof. Verhasselt for their invaluable discussion on handling missing data, as well as Dr. Vercruysse and Mrs. Machiels for their administrative support throughout this research and Mr. Vandebempt for his invaluable assistance with paperwork that made my enrollment in this Master's program possible.

Abstract

Multiple Sclerosis (MS) is an incurable chronic disease that affects the conduction of neurons in the central nervous system. Evoked potentials (EP) are measurements that give insight into the damage caused to these neural pathways. Motor evoked potentials (MEP) specifically measure the level of degradation of signals transmitted from the brain to muscles in the hands and feet. Preliminary work in the research group of Peeters et al. has proven that evoked potential features have predictive value when predicting MS disease progression (Yperman et al., 2020a), (Yperman et al., 2020b). However, it is clear that the trajectory of evoked potential features greatly differs between individuals. Therefore, it is difficult to identify clinically meaningful deterioration of MEP over time for a specific individual. Hence, there is a need to identify individual reference intervals (IRI) for MS biomarkers, specifically focusing on longitudinal MEP measurement. The research group of Thas et al. has recently developed an innovative methodology to calculate IRI in a real-world setting and provided proof-of-principle using open data (Pusparum et al., 2020), (Pusparum et al., 2021). The IRI takes into account someone's specific characteristics reflected by his/her past data, together with peer's data from the same population, providing a personalized and more precise interpretation of results.

An open-source dataset that included 693 people with MS (PwMS) with MEP measurement (Yperman et al., 2022) is used. From then, 98 PwMS were chosen to perform IRIs. It was found that IRIs have a high accuracy (minimum of 60% and maximum of 88 %) of detecting the progression in MS, with 16 months on average before EDSS score detects the progression. Although the results are different between the arm/leg (71.5% versus 72%) and the left/right side datasets (62% versus 81.5%), the overall conclusion is consistent.

Our study suggests using IRI on MEP measurement as a support method in diagnosing the progression of MS. However, the application of our result should be treated carefully, with attention to the small sample size that is used to perform IRI.

Keywords: *Multiple Sclerosis, Individual Reference Interval, Motor Evoke Potential, EDSS score*

Contents

1 Introduction

1.1 Multiple Sclerosis

Multiple Sclerosis (MS) is a chronic neurological autoimmune disease that impacts an estimated 2.9 million individuals globally, with females more than males (Atlas), (Haki et al., 2024). MS is characterized by the immune system mistakenly attacking the protective covering of nerve fibers (myelin) in the central nervous system and can lead to a wide range of symptoms, including physical, mental, and sometimes psychiatric problems, decrease the overall quality of life (WHO). Figure 1 presents the MS rate across the world. MS rates are higher in high-income countries, particularly Europe and the United States than in other regions. Specifically, Belgium has a prevalence rate of 104 per 10000 people, which is relatively high compared to the global average. This statistic highlights the critical importance of MS within the context of Belgium's healthcare landscape, underscoring the necessity for sustained research, and advancements in interventions in order to support individuals grappling with this complex disorder (Atlas).

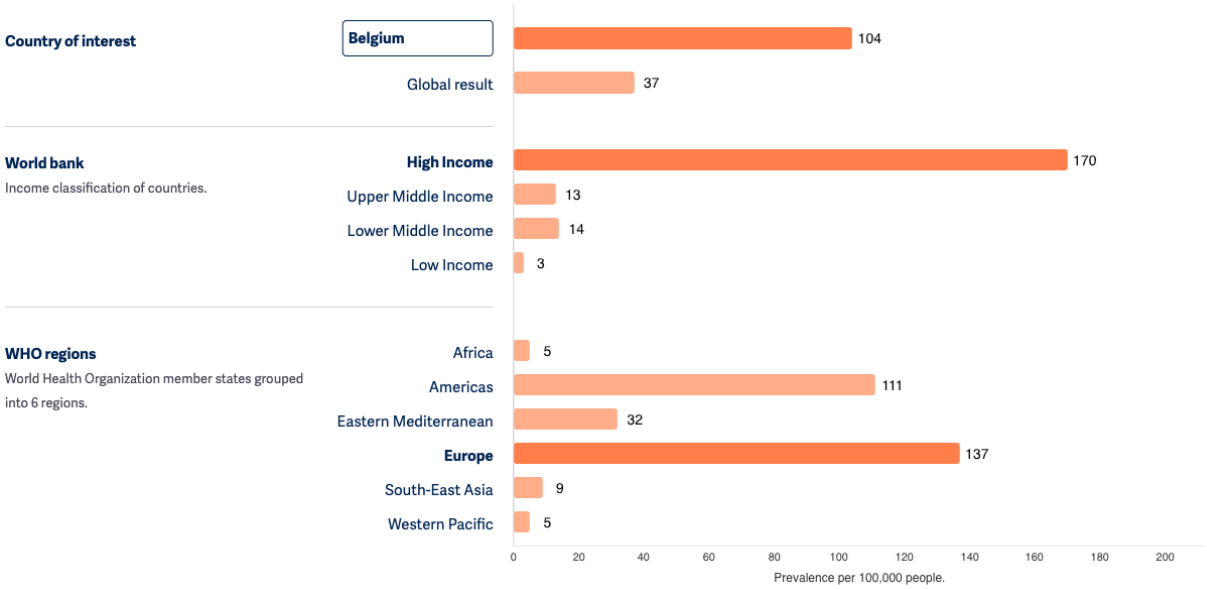


Figure 1: Prevalence per 100000 people of MS in Belgium and worldwide. The darker the color, the higher the prevalence (Atlas).

The majority of individuals diagnosed with MS have the relapsing-remitting type (RRMS), marked by periods of new or worsening symptoms (relapses) followed by periods of recovery (remission). Over time, RRMS may evolve into secondary progressive MS (SPMS), where the disease progresses more steadily, with or without periods of remission. Conversely, a smaller subset of PwMS is diagnosed with primary progressive MS (PPMS) from the outset,

marked by a gradual worsening of symptoms without distinct relapses and remissions (riv, 2022). MS can be diagnosed by showing the dissemination of lesions in space and time. In practice, magnetic resonance imaging of the central nervous system, along with the oligoclonal bands, is used to support the diagnosis (Polman et al., 2011), (Milo and Miller, 2014), (Freedman et al., 2020). The expanded disability status scale (EDSS score) is often used to assess the progression of a PwMS's condition (Fischer et al.), (Cutter, 1999). EDSS score was first defined in 1983 (Kurtzke, 1983) and is widely accepted in clinical trials. However, besides many good characteristics such as objectivity, reliability, and validity, EDSS score is not sensitive to change, making the timely diagnosis of progression challenging. (Kragt et al., 2006), (Noseworthy et al., 1990). Besides EDSS score, neurophysiological techniques, such as MEP, are often recommended in the context of MS (Fernández et al., 2013).

MEP serves as a crucial measure for monitoring the progression of MS. They potentially provide a significant advantage over EDSS score by detecting changes in the condition more sensitively and rapidly (Hardmeier et al., 2017), (Hardmeier et al., 2020). MEP measure the electrical activity in response to nervous system stimulation, offering real-time insight into the functional status of neural pathways. This method is especially effective in tracking the progression of MS, as it captures subtle changes in neurological function that might not yet be apparent through clinical assessments like the EDSS score. Given its higher sensitivity and faster response to neurological changes, MEP is increasingly recognized as a valuable tool in both research and clinical management of MS (Leocani et al., 2000), (London et al., 2017), (Yperman et al., 2022).

1.2 Mutiple Sclerosis in personalized medicine

Disease-modifying therapy (DMT) for MS aims to reduce the time to progression and reduce the frequency and severity of relapse, hence improving the overall life quality. DMT can be different for each individual, depending on the condition and type of MS (Robertson and Moreo), (Gold et al., 2010), (Freedman et al., 2020). There is no DMT that fits all the PwMS. Therefore, tracking the PwMS's response and adjusting the management plan in time is crucial (Giovannoni and Rhoades, 2012), (Apóstolos et al., 2022), (Marriott et al., 2023).

The reference interval is a range of values derived from a clinical test. Healthy reference intervals are typically calculated based on a healthy population. If a patient takes a laboratory test for a particular clinical biochemistry marker, and the result falls within this range, it indicates good health. However, if the result is outside of this range, further testing may be necessary to evaluate the patient's condition (Pusparum et al., 2021). While the reference interval is commonly used, it only gives population interpretation, hence the same range for individuals. When there is a significant variability in laboratory test results among individuals with a particular condition, the reference interval may become insen-

sitive. For the context of MS, a clear recommendation for the reference interval of MEP cannot be found in the literature. This research uses the reference interval mentioned in (J., 2007), which is accepted in practice.

To tackle the disadvantage of reference interval, individual reference interval (IRI) will be used. It specifies an interval for each individual, helping to detect the problem as soon as possible. The approach to health care based on individual care is called personalized medicine (Stefanicka-Wojtas and Kurpas, 2023). In the case of MS, personalized medicine can give many benefits and help clinicians diagnose faster and better. To interpret MEP accurately, it is essential to establish the normal range of MEP values for each individual. As MEP is measured on PwMS, a normal range refers to the range in which one is in a stable period and does not experience any disease progression. This step is crucial to accurately assess what constitutes a normal versus an abnormal MEP reading for each PwMS. Understanding the normal range will allow for more precise evaluations of the disease.

The algorithm developed by Thas and Pusparum introduces a novel approach to nonparametric estimation, significantly reducing the data requirement to just a few points per individual. This innovative method facilitates more efficient and accurate analysis in situations where data collection is challenging, offering a practical solution for researchers and practitioners in fields where detailed individual datasets are too few or difficult to obtain (Pusparum et al., 2020), (Pusparum et al., 2021).

This study aims to evaluate the effectiveness of using IRI on MEP to detect the progression of MS, and compares its sensitivity and accuracy to that of the established EDSS score scoring system. This thesis is organized as follows: Section 2 focuses on the methodology that is used to run and evaluate the IRI, along with the description of the open-source dataset. Two main measurements used in constructing IRI in the context of MS are EDSS score and MEP, which are also discussed in more detail. Section 3 describes the study design, focusing on data preparation for running IRI. Finally, Section 4 presents conclusions and remarks in a clinical context. The Appendix A - Appendix O show the results of all datasets and the selected code used for this study.

2 Methodology and Data

In this section, the principal methods to find reliable IRIs are described. Along with this is the summary of the open-source dataset (Yperman et al., 2022).

2.1 Methodology

The IRI is estimated using a non-parametric method called penalized joint quantile model (Pusparum et al., 2021). The 95% confidence interval (CI) of the IRI is calculated using the Bootstrap, with 500 replications each time. Four subsets are used: the right hand, the right

foot, the left hand, and the left foot datasets. The division is based on the assumption that the side and the limbs are independent. The two measurements for MS are EDSS score and MEP. This section briefly describes how they are defined and calculated, along with their advantages and disadvantages, and how that information is used to study the potential new method to detect the progression of MS.

2.1.1 Reference Interval Estimation

A quantile mixed-effect model is applied to estimate the IRI. A quantile is the inverse of a cumulative distribution function

$$q_\tau = F_Y^{-1}(\tau) = \inf\{y : F(y) \geq \tau\},$$

where

- τ is the probability, $\tau \in [0, 1]$.
- F^{-1} is the inverse cumulative distribution function.

Penalized quantile mixed model

The joint quantile mixed model is defined as

$$\begin{aligned} Q_i(0.5) &= \beta_0 + u_i, \\ Q_i(\tau_1) &= \beta_0 + u_i + z_i\beta_1, \\ Q_i(\tau_2) &= \beta_0 + u_i + z_i\beta_2, \end{aligned}$$

where

- $Q_i(0.5)$ is the 50 percentile of the *i-th* patient.
- $Q_i(\tau_1)$ ($Q_i(\tau_2)$) is the τ_1 (τ_2) percentile of the *i-th* patient. It is often chosen such that $\tau_1 < \tau_2$ and refers to the lower and upper boundaries of IRI of the *i-th* patient.
- β_0 is the fixed intercept, referring to the population level.
- u_i, z_i are the individual specific coefficients.

The individual interval is estimated by the difference between the τ_1 percentile and τ_2 percentile. This kind of model allows us to adapt the quantile value and the length of the individual interval. It is derived from the fact that even when different models are estimated, they originate from the same person and thus must share some parameters.

The parameter β_0 is estimated by the median of all data. The parameters β_k, $k = 1, 2$ and u_i, z_i with $i = 1, ..., N; j = 1, ..., n_i$, are estimated by minimizing the function

$$M(\beta_1, \beta_2, u, z; \lambda) = \sum_{k=1}^{2} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \rho_{\tau_k}(y_{ij} - \beta_0 - u_i - z_i\beta_k) + \lambda_u \sum_{i=1}^{N} u_i^2 + \lambda_z \sum_{i=1}^{N} (z_i - 1)^2,$$

where

- $\rho_\tau(w) = w(\tau - w \leq 0)$ is the check function.
- The first term aims to minimize the objective function of the linear quantile model.
- The second and third terms are the penalized terms. The choice of λ_u and λ_z determines the variance between the subject-specific median and the variance between the IRIs.

Estimation procedure

To estimate the parameters, two steps are followed

- Step 1: Estimate β_0 by the median of all data and initial estimation of u_i, β_1 and β_2.
- Step 2: Until convergence, do the loop.
  - Estimate z_i.
  - Estimate β_1 and β_2.
  - Estimate u_i.

Selection of penalty parameters

The penalty parameters (λ_u and λ_z) are chosen to optimize the coverage probabilities. The coverage probability is the probability that the reference interval will cover the outcome. This method uses two coverage probabilities: Time Empirical Coverage (TEC) and Subject Empirical Coverage (SEC). The former refers to the probability of coverage for a new measurement of a subject, which is included in the stable data for constructing the IRI. The latter refers to the probability coverage for a new measurement of a new subject, which is not included in the stable data for constructing the IRI.

Confidence intervals for IRI

As the penalized joint quantile model is non-parametric, non-parametric Bootstrap is chosen to calculate the 95% CI of IRI. Bootstrap creates a large number of resampled data based on the original dataset. In each sample, the parameters (lower and upper boundaries for each PwMS) are calculated based on the algorithm 2.1.1, and the distribution of parameters is given. In this study, the original dataset was resampled with replacement 500 times. Then, the 2.5th, 50th, and 97.2.5th percentiles of these 500 values were performed. If the estimated IRI from the original dataset is inside the 95% CI, ideally close to the median of the 500-value set, it is considered that the IRI is well estimated.

2.1.2 EDSS score and MEP

EDSS score scale assesses disability through neurological examinations, focusing on eight principal functional systems: Pyramidal, Cerebellar, Brain Stem, Sensory, Bowel and Bladder, Visual, Cerebral (Mental), and Other (Miscellaneous). Each system is rated on a 0 (normal) to 5 or 6 (severe impairment) scale. The "Other" category scores 1 for any impairment and 0 for none. Kurtzke stressed not combining scores but comparing them independently over time to monitor changes. Despite its heavy dependence on functional systems for lower scores, EDSS score is closely aligned with leg mobility. For example, EDSS score of 3.0, 3.5, and 4.0 are characterized by full ambulation, which means they can be up and about all data and walk a usual distance without resting. A summary of the EDSS score is shown in Figure 2.

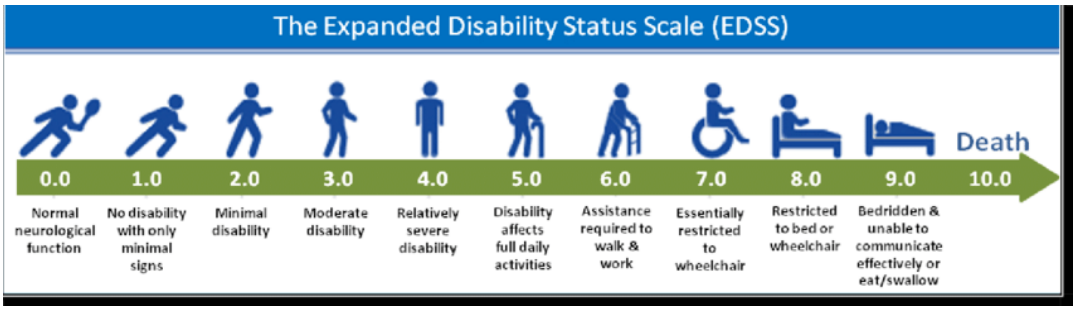


Figure 2: Illustration of PwMS's condition corresponds with EDSS score with one unit increasing each time. Source: MS website.

EDSS score is extensively utilized in clinical trials due to its strengths in objectivity, reliability, and validity. Despite its wide acceptance, the approach to interpreting the EDSS score remains undefined. In this thesis, the guidelines proposed by (De Brouwer et al., 2021) have been followed to assess progression in individuals with MS over a period of two years. Clinically, this timeframe is regarded as essential to identify progression, considering that a shorter duration might overlook gradual yet continuous increases in EDSS scores. It is

critical to emphasize the significance of the selected time interval, as every person with MS is expected to experience disease progression at some point in their lifetime. A progression in MS happens if one of these conditions is satisfied:

- An increase of at least 1.5 if $EDSS = 0$.
- An increase of at least 1 if $0 < EDSS \leq 5.5$.
- An increase of at least 0.5 if $EDSS > 5.5$.

Another drawback of EDSS score is that it is not sensitive to change (Hutchinson and Hutchinson, 1995), (Hohol et al., 1999), (Meyer-Moock et al., 2014), so the moment that a progression of MS due to the change in EDSS score is detected, it may be too late to adapt the DMT. **This motivates us to find another scale or a support scale to measure the progression of MS.** There are some alternative scales to EDSS score, such as The Ambulance Index (Hauser et al., 1983), the Scripps Neurological Rating Scale (Sipe et al., 1984), the Guy's Neurological Disability Scale (Sharrack and Hughes), the Multiple Sclerosis Impairment Scale (Ravnborgl et al., 1997). However, each of these instruments comes with limitations. Hence, this study focuses on a neurophysiological measurement: MEP.

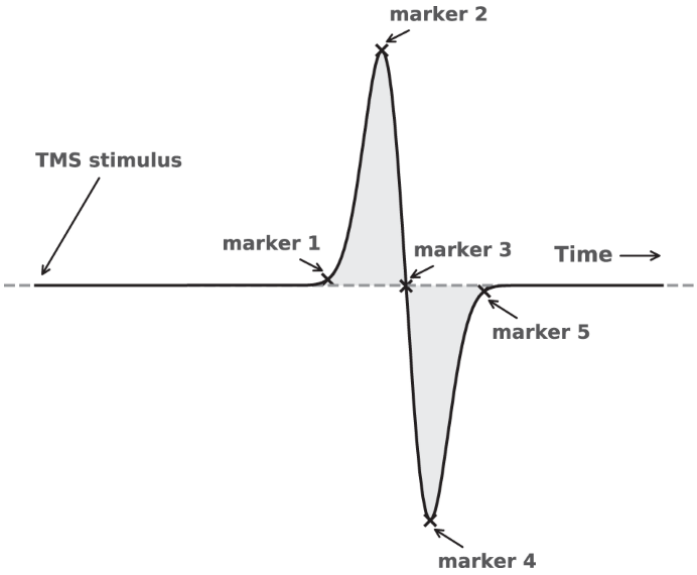


Figure 3: A simplified trajectory of MEP measurement. TMS stands for Transcranial Magnetic Stimulation. There are five markers in total. The most important one is marker 1 (Yperman et al., 2022)

.

EP is a clinical test that measures electrical signaling through the motor pathway of the nervous system. In PwMS, this signal is decreased due to the damage in demyelination. To

perform MEP, the motor cortex is excited, and the signal to the feet or hand is measured. Change in MEP in a longitudinal study is shown to be correlated to the progression in MS (Leocani et al., 2000). With the help of technology nowadays, these tests are easy to perform. However, the EP measurement often has high variability due to the instability in conduction in the damaged axons and/or bad cooperation of the PwMS. A typical MEP measurement is shown in Figure 3. There are five markers in total. The first marker is the moment the signal leaves the x-axis. The second marker is the highest peak of the signal, the third marker indicates when it crosses the x-axis, the fourth marker is the lowest point, and the last marker is when the signal returns to zero. As the signal can have great fluctuation, the first marker (marker 1) and the difference between the second and the fourth markers (marker 4 - marker 2) are often used to detect the progression in MS (Yperman et al., 2022).

2.2 Data

An open-source data which contains 693 PwMS observed during six years (Yperman et al., 2022) are used. The results of this study will be used to support clinical practice and improve individual DMT, hence benefiting PwMS. There are five main datasets: edss.csv, measurement.csv, patient.csv, and test.csv. The details of each dataset and how to merge them together to run IRI's algorithm will be explained. The overview of the datasets is shown in Figure 4.

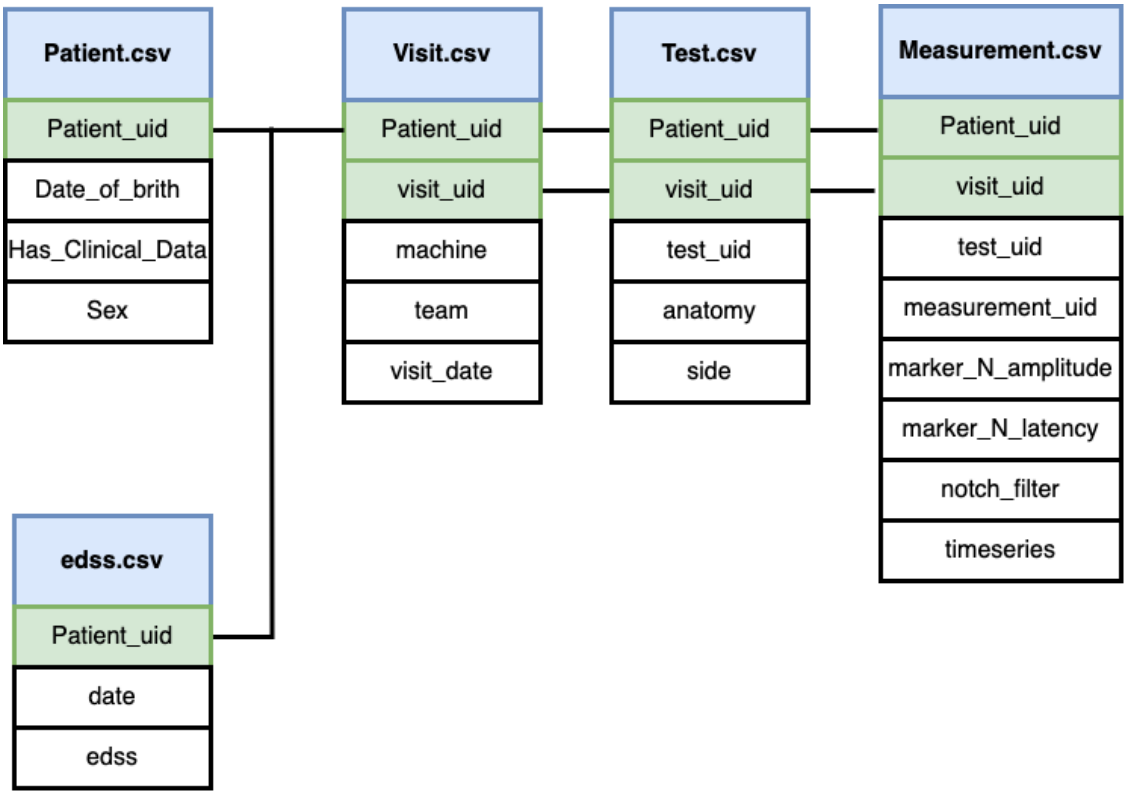
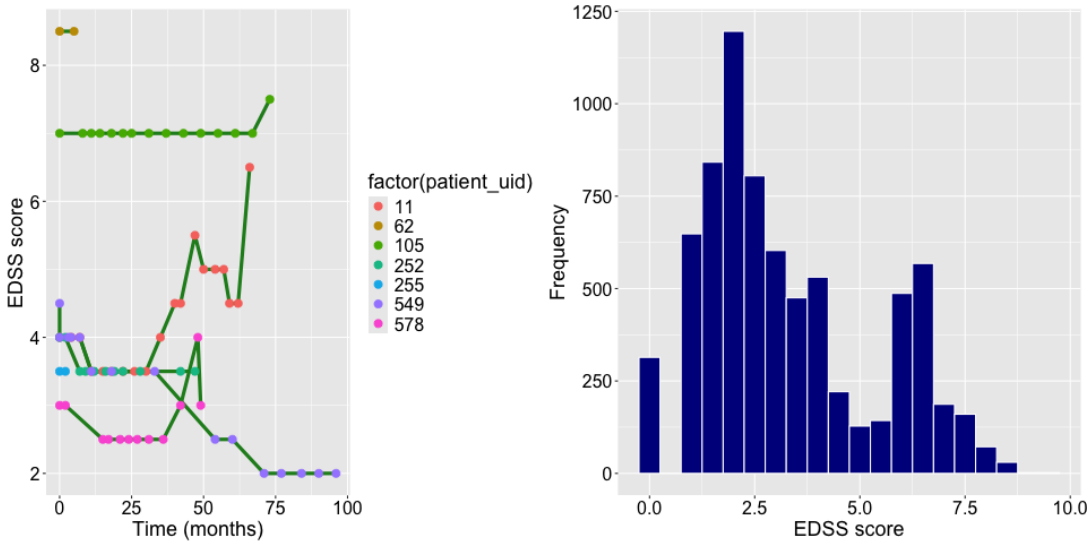


Figure 4: An overview of five datasets, the blue box refers to the name of each dataset, and the green boxes are the same information (variables) shared among the datasets. The black lines connect different datasets together using common variables.

- ***edss.csv*** is longitudinal, containing the information about the EDSS score of each PwMS over time.
  - *date*: It is the date that the EDSS score is measured.
  - *edss*: The value of EDSS score on that exam day.
  - *patient_uid*: It is the unique identifier for the PwMS. There are, in total, 582 PwMS with recorded EDSS score values measured at several visit times on different dates.



(a) (b)

Figure 5: (a) Examples of the EDSS score evolution in seven PwMS, randomly selected. The trajectories have numerous patterns. Some PwMS do not experience progression during the observed time. (b) The histogram of EDSS score of all PwMS (582) on the right. The highest frequency of EDSS score is around the point 2.5.

Figure 5 shows the EDSS score revolution of seven randomly chosen PwMS on the left. The difference between time visits is calculated using month unit. The origin is the first time the PwMS has recorded EDSS score. It is observed that the pattern of EDSS score varies a lot between PwMS, and it is a non-monotone pattern. The number of time points also differs significantly. On the right is the histogram of 582 PwMS. It is observable that most PwMS have EDSS scores in a low range from one to three.

Since our goal is to study the evolution of MS, not all the EDSS score information for each PwMS is utilized. The time gap between visits varies from three months to two years, depending on the individual. A common approach in the literature is to verify the change in EDSS score every two years to ensure that it does not overlook the progression characterized by a slow but steady increase in EDSS score over time. For example, if a PwMS's EDSS score increases by 0.25 points every six months, it would fail to detect the progression if it only considers the difference in EDSS score between two consecutive visits. However, by examining the difference in EDSS score with a two-year interval, a difference of one point can be calculated, confirming progression in MS.

- ***Measurement.csv*** contains MEP information. The most important variable that

needs to pay attention to is *marker*1_*amplitude.mv*. Hereafter, it is called *latency* measurement. It is the time the signal is needed to reach an individual's arm or foot.

  - *marker_N_amplitude.mv.*: It refers to the amplitude of the marker. N takes the value in 1, 2, 3, 4, 5 corresponding to the five markers. Their position can be seen in Figure 3. The smaller the amplitude, the more severe the PwMS's situation.
  - *marker_N_latency.ms.*: It refers to the latency of the marker, N takes the value in 1, 2, 3, 4, 5. The higher the value, the more severe the PwMS's situation.
  - *measurement_uid*: It is the unique identifier for the measurement.
  - *notch_filter*: It is a true/false variable indicating whether a notch filter was applied during the measurement.
  - *test_uid*: It is the unique identifier for the different tests within a visit.
  - *visit_uid*: It is the unique identifier for the visit of a PwMS.
  - *time_series*: It is the time series of each PwMS at each measurement.

- ***Patient.csv*** describes some extra information of PwMS. It includes four variables:
  - *date_of_birth*: It is the birthday of each patient.
  - *has_EDSSscore_measurements*: It is a true/false variable, indicating whether a PwMS has EDSS score.
  - *sex*: It is the gender of each individual.
  - Age and gender can used as covariates in modeling.

- ***Test.csv*** gives some extra information about MEP measurement.
  - *anatomy*: It is the limb that is measured; there are two values, APB and AH. APB refer to the arm, and AH refer to the leg.
  - *side*: It is the side of the body that is measured; there are two values L for the left side and R for the right side.

- ***Visit.csv*** gives the information relating to each visit of each PwMS
  - *visit_date*: It is the date of each visit.
  - *machine*: The machine was used during that visit. There are two machines: A and B.
  - *team*: The technicians performed the test during a visit. There are two teams, A and B.

3 Study Design

The study's design is aimed at preparing the data for the performance of IRI. The concept of IRI, which is based on healthy individuals, ideally requires data from individuals before they are diagnosed with MS, which is unattainable. Data are only available from individuals after they have been diagnosed with MS. Therefore, to implement the IRI concept, stable data from each person with MS are chosen for the interval analysis. Whether a person with MS is classified as stable or unstable is determined by the gold standard EDSS score. The interval is used to compare future data in cases where progression has been experienced by a person with MS, in order to study the effectiveness of IRI. Figure 6 outlines the study design chosen to address the study's questions. The design consists of four main steps:

- Step 1: Prepare data to find the *breakpoint*.
- Step 2: Find the *breakpoint* and prepare data to check for stability.
- Step 3: Check the stability of latency, prepare data to run IRI
- Step 4: Calculating IRI, perform Bootstrap for evaluation IRI.

It is worth noting that two study designs are followed in this study. The first does not consider the third step, whereas the second follows the four steps above. The decision to conduct the IRI analysis without the latency stability check was driven by two primary considerations aimed at understanding the sensitivity of the IRI to certain conditions and the potential early indicators of disease progression not captured by the traditional instrument.

The first reason is to assess the IRI's sensitivity to unstable data. This approach allows us to explore the robustness of the IRI when faced with fluctuations in the data, providing insight into how such instability could impact the interpretation of results. From then on, suggestions on how to manage the data can be given.

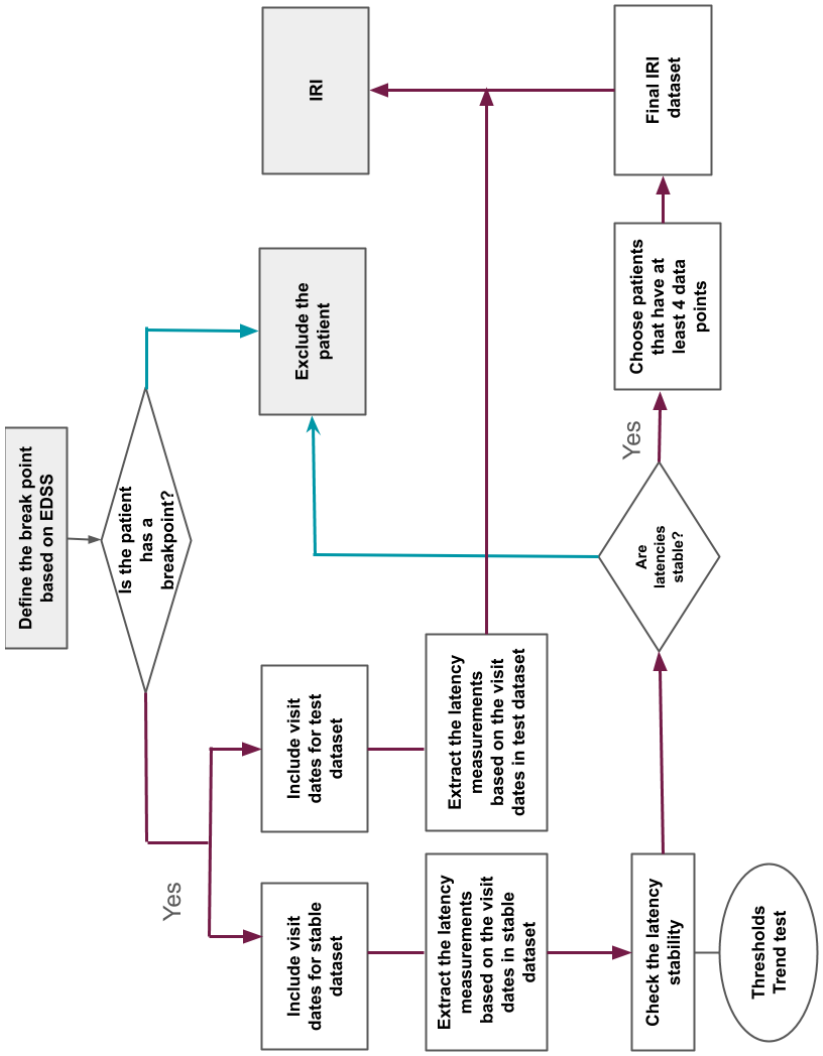


Figure 6: Study Design for this study from the second step to the fourth step. The red lines represent how stable data is chosen for running IRI. The blue lines present how the PwMS who do not satisfy certain conditions are excluded from the stable data.

The second reason comes from the fact that the potential of latency measurements as a more immediate reflection of disease progression compared EDSS score. By bypassing the latency stability check, it aims to identify whether individuals who are otherwise excluded from the dataset due to not meeting certain threshold criteria might be exhibiting early signs of progression not yet evident in their EDSS scores. It can be seen as an important signal of considering alternative or supplementary measurements to the EDSS score for a comprehensive PwMS status and disease evolution evaluation. Hereafter, it focuses on explaining the study design, which includes the four steps; for the other design, it simply skips the third step and jumps directly to the fourth step from the second step.

To begin, it needs to prepare the data to identify *breakpoint*. This first step is not shown in Figure 6. *breakpoint* is a date that divides the data into two parts: the period from the first hospital visit up to this date is considered to reflect stable EDSS scores. Unstable EDSS score evolution, which leads to progression in MS, is defined in Section 2.1.2. It is important to note that the EDSS score used for evaluation is not the original value in ***edss.csv***. It has already been transformed to reflect a time interval of two years. Initially, all PwMS for whom latency measurements were available were included. Individuals without this score are excluded since *breakpoint* is determined based on the EDSS score. Furthermore, data for PwMS without age and sex information are also excluded. To calculate the difference in EDSS score, at least two visits are required, so PwMS with only one EDSS score visit are also excluded.

The second step is shown at the beginning of Figure 6, with the gray rectangle: "Define the breakpoint based on EDSS." After identifying whether a PwMS has a *breakpoint*, only those who have experienced progression at least one time during the observed time are kept for the next step. Then, following the red lines, the latency data was separated into two categories: stable latency (stable dataset) and future latency (test dataset). Since *breakpoint* is determined based on the EDSS score and applied to latency data, there are cases where no future data is available for some individuals. Those individuals were excluded as the aim was to assess the performance of IRI using future data. If the future data is outside the IRI, it indicates that the latency value has become unstable, possibly due to an upcoming progression.

The third step starts at the diamond box located at the bottom of Figure 6, written: "Are latencies stable." In this step, latency stability within the stable data is verified. Reminding that "stable data" refers to the latency data during a period of time in which the EDSS score is stable, whereas "latency stability" refers to the latency measurement itself, that it does not contain outliers or increasing trend. After this verification, the chosen data are utilized to conduct IRI. This step is necessary since the evolution of latency and the EDSS score may be different. For this verification, two sub-steps are followed. Initially, latency measurements are checked against a normal range, set at [34.4 - 48] (ms) for the

foot and [18.7 - 23.5] (ms) for the hand (J., 2007). Measurements falling outside of this range are considered abnormal. Secondly, a trend in the stable data is detected through the execution of a modified Mann-Kendall test, employing the R package modifiedmk. The test hypothesis is:

$$H0 : \text{There is no trend}$$
$$H1 : \text{There is a trend}$$

A significant level of 0.05 is used, if the test is significant, the individual is excluded out of the data.

In the last step, the IRI algorithm is performed for each subset, and the effectiveness of IRI in detecting the progression is verified by calculating the accuracy

$$\text{Accuracy} = \frac{\text{The number of visits whose latency is close to extremes or outside the IRI}}{\text{The total number of visits}}$$

As in our study design, only the PwMS who experience at least one progression is included inside the dataset, the accuracy can be understood as sensitivity of the latency using IRI. The specificity cannot be calculated as all the PwMS who do not experience progression are left out.

Finally, a Bootstrap with 500 replications is performed to calculate the 95% CI of the IRI. With this result, the exact time that latency using IRI can detect the progression is calculated and compared to the time to detect the progression using EDSS score.

4 Results

4.1 Data management

The datasets are merged and cleaned to be ready to find *breakpoint* and test latency stability. From 693 PwMS from the largest datasets (measurement.csv), of which 98 PwMS were included for testing the latency stability, 595 PwMS were excluded based on different requirements. The result is summarised in Figure 7.

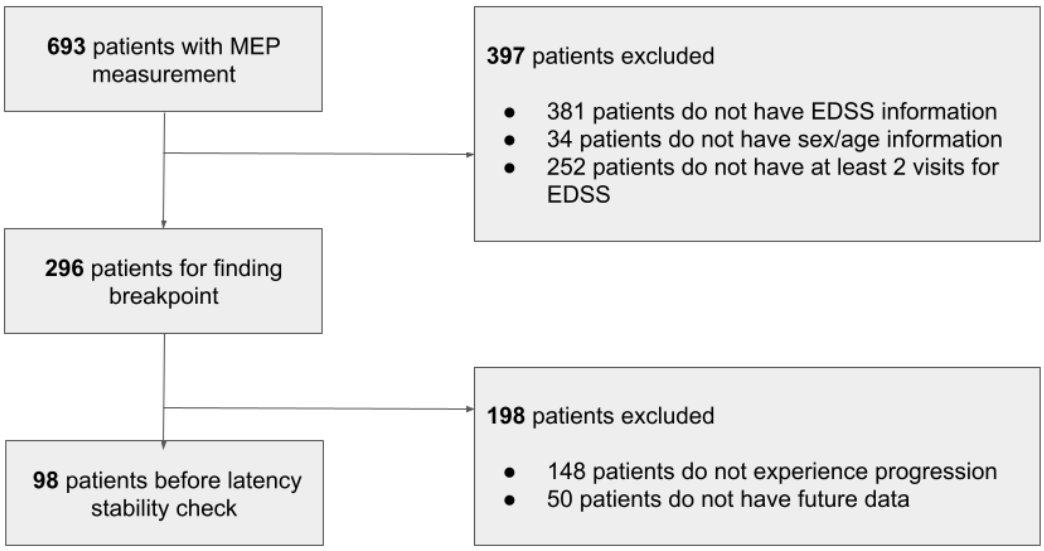


Figure 7: Flow diagram detailing the selection process of PwMS included for latency stability check (steps 1 and 2 in study design)

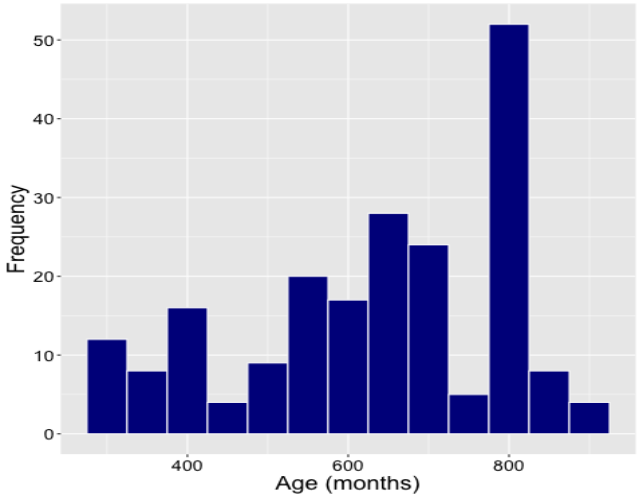
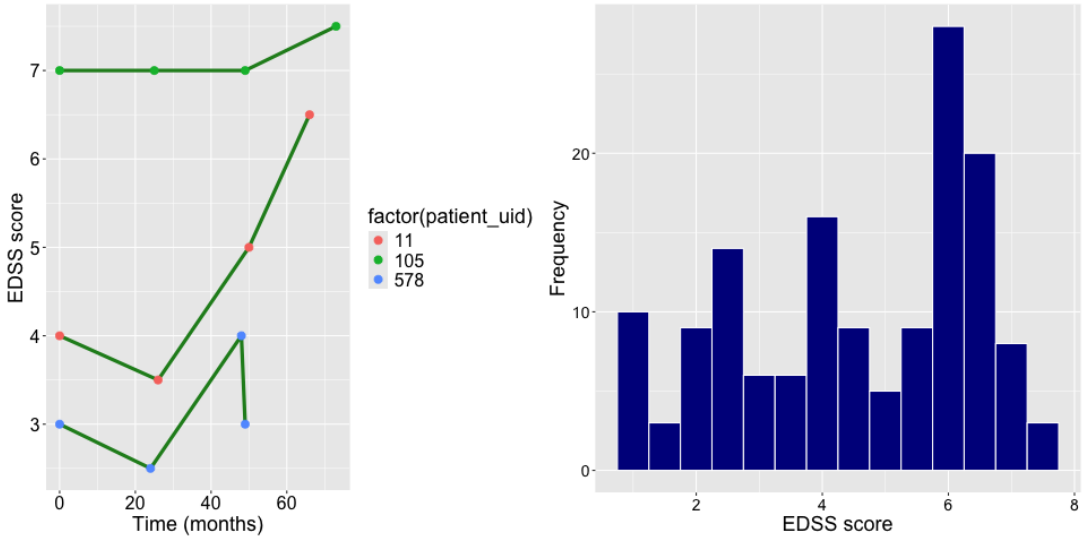


Figure 8: Age distribution of 98 individuals with MS before the latency stability check. Age indicates the age at the first hospital visit. The most frequent ages are 50 and 67 years old.

After the first and second steps in study design, there are 98 PwMS. Out of these, 66 are female and 32 are male. The age distribution of these 98 PwMS is shown in Figure 8. The

highest frequency is observed at 800 months, which is equivalent to 67 years old. This is the age of PwMS at their first hospital visit. In Figure 9 on the right, the histogram of EDSS score is displayed for these 98 PwMS. The histogram still shows a bimodal distribution. Unlike the previous EDSS score histogram, the most frequent EDSS score value now falls around 6-6.5 rather than 2-2.5. Most of the PwMS who experience the progression in MS shift to a higher EDSS score. It may be due to the nature of the disease, where RRMS PwMS with stable EDSS score for years will eventually jump into SPMS, with a steady development in the symptoms, including an increase in the EDSS score. Although the decision to use two years to determine whether a PwMS experiences progression is widely clinically accepted, it can have some limitations in tracking the evolution of MS in the early period. The same seven random PwMS are chosen on the left side of the same figure to observe the EDSS score evolution as in Figure 5. However, after the data management, four PwMS are excluded, only three stayings (uid 11, uid 125, uid 678) with *breakpoint* at 50, 75, and 48 months after the first visit for checking EDSS score.



(a) (b)

Figure 9: (a) Examples of the EDSS score evolution in three PwMS, randomly selected. The time interval between two sequential EDSS scores is 24 months. All PwMS experience progression once during the observed time. (b) Histogram of EDSS score of 98 PwMS (on the right), the highest frequency of EDSS scores is 6.5.

From 98 PwMS who have at least one progression in the observing time, four sub-datasets are made: right hand, right foot, left hand, left foot. It is based on the information in *anatomy* and *side* in the ***test.csv*** dataset. The details of PwMS included criteria in each

step when considering latency stability are shown in Table 1 and when not in Table 2. It is trivial that the number of PwMS in the former is smaller than in the latter.

Table 1: Details the selection process of PwMS included for the IRI algorithm, corresponding to step 3 in the study design. The column *Data* shows the number of PwMS in each subset after steps 1 and 2 in the study design. The column *Latency stability* shows the number of PwMS in each subset after checking latency stability using clinical threshold. The column *Minimum 4 data points* shows the number of PwMS in each subset after choosing the individuals who have at least 4 data points before *breakpoint* date. The last column, *Trend test*, shows the number of PwMS in each subset after the last check for stability using a trend test for longitudinal data.

	Data	Latency Stability	Minimum 4 data points	Trend Test
Right Foot	90	73	33	31
Left Foot	89	74	33	33
Right Hand	96	78	35	34
Left Hand	97	72	31	29

Table 2: Details the selection process of PwMS included for the IRI algorithm, corresponding to step 3 in the study design. The column *Data* shows the number of PwMS in each subset after steps 1 and 2 in the study design. No latency stability check is performed. Only a selection for a minimum of 4 data points is made. The column *Minimum 4 data points* shows the number of PwMS in each subset after choosing the individuals who have at least 4 data points before *breakpoint* date.

	Data	Minimum 4 data points
Right Foot	90	43
Left Foot	89	43
Right Hand	96	48
Left Hand	97	48

4.2 Individual Reference Interval for latency

After the data management, four subsets are divided. This subsection shows the result of IRI on the latency measurement, on the right hand dataset, and the right foot dataset. The results for the left hand dataset and for the left foot dataset are shown in Appendix. There are in total three different IRIs are shown for each dataset:

- IRIs without a latency stability check.

- IRIs with a latency stability check.
- IRIs with latency stability check and using the information sex and age.

IRI without latency stability check

Figure 11 and Figure 12 show the IRI when there is no control on latency stability. IRI detects the progression when there is at least one future value (colored round points) that is close to the extremes or outside of the interval. For the right hand data, 48 PwMS are included in calculating IRIs; 8 PwMS are excluded in calculating the accuracy as they have only one future point, which is considered too few. Finally, 18/40 (45%) PwMS with more than one future point can be detected with progression using IRI. For the right foot data, 43 PwMS are included in calculating IRI, and 7 PwMS are excluded when calculating accuracy. Finally, 20/36 (55%) PwMS with more than one future point can be detected with progression using IRI.

From these two graphs, it is observed that the IRIs are, in general, very wide. The largest interval for the foot is around [15, 50] (ms), knowing that the population interval is around [34, 48] (ms) (J., 2007). The largest interval for the hand is around [15, 50] (ms), knowing that the population interval is around [18, 24] (ms) (J., 2007). Therefore, a low detection ability for progression is expected. The wide IRI can be explained firstly by extreme values that appeared in stable data in some PwMS, for example, id 271 and id 460 for the right hand data; id 615 and id 946 for the right foot dataset. Secondly, for some PwMS, the whole stable data is generally high, for example, id 188 and id 280 for the hand right data; id 449 and id 552 for the right foot data. This may be attributed to the fact that the progression of MS is potentially captured faster by latency than by EDSS score. Consequently, PwMS who have already undergone progression were included in this case, and the IRIs no longer represent the patient's healthy status.

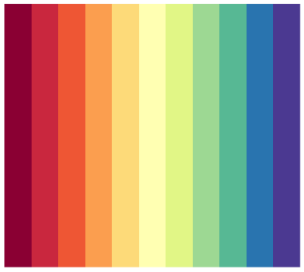


Figure 10: The color range for the future data. From left to right corresponds to the time the future data is recorded. The closest to *breakpoint* is red, and the furthest point to *breakpoint* is violet. This color code is applied to all the graphs in this study.

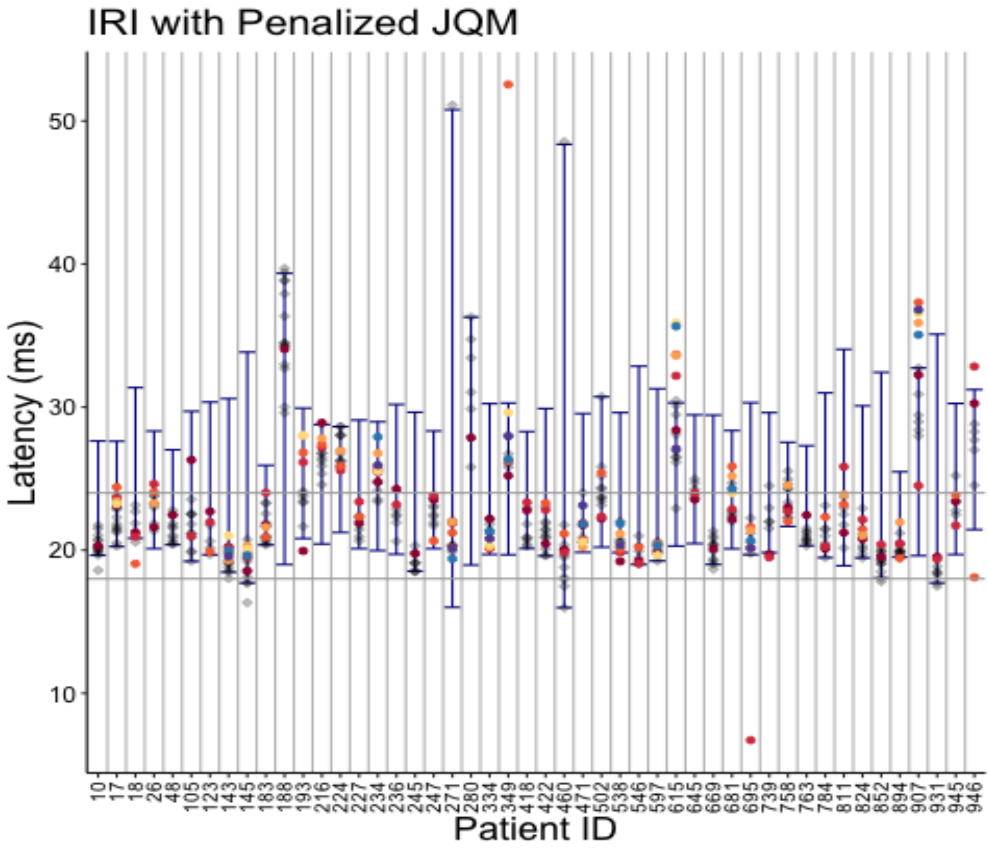


Figure 11: IRI from the right hand data, without latency stability check. The blue intervals are the IRI. The gray diamonds are the stable data. The colored points are future data. The more blueish the point, the further in the future the data is. The two horizontal gray lines represent the population reference interval.

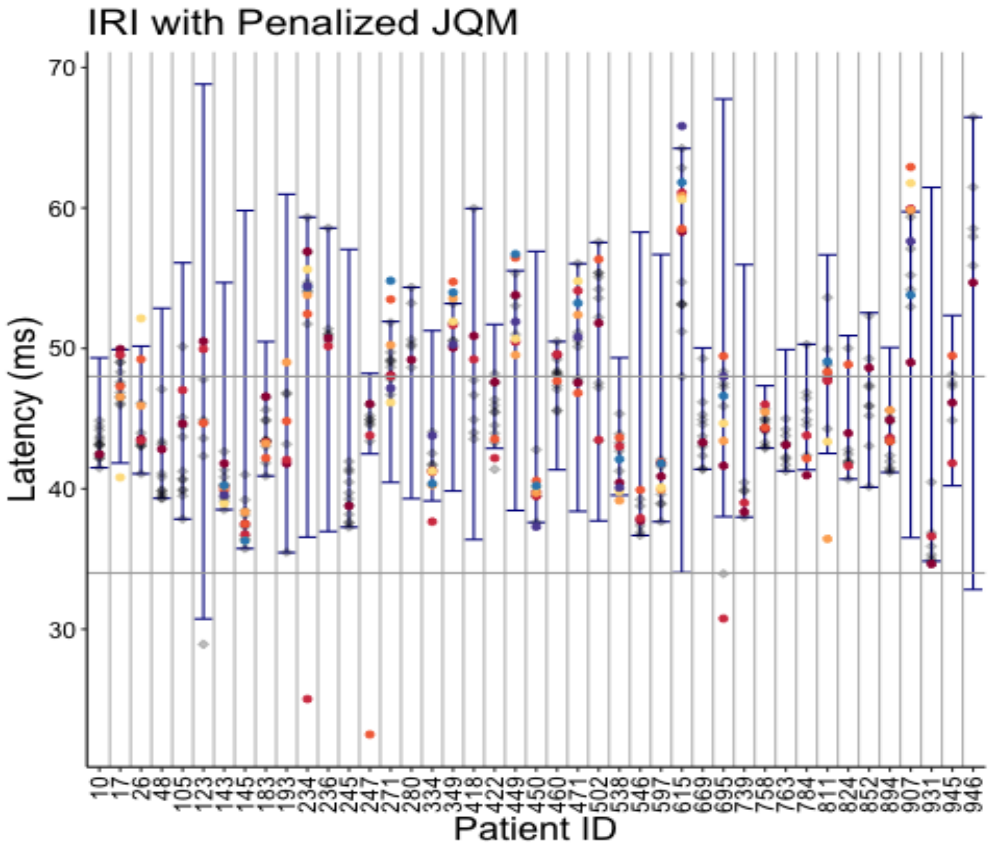


Figure 12: IRI from the right foot data, without latency stability check. The blue intervals are the IRI. The gray diamonds are the stable data. The colored points are future data. The more blueish the point, the further in the future the data is. The two horizontal gray lines represent the population reference interval.

IRI with latency stability check

Figure 13 shows the IRI of 34 PwMS from the right hand data. It is observed that 23/29 (80%) PwMS with more than one future point can be detected with the progression using IRI. There are 17/29 (59%) that the future points are close or greater than the upper bound of the interval, and 6/29 (20%) that the future points are close or smaller than the lower bound of the interval. In the first case, it indicates that the future latency is getting larger than the normal range; it hints that the time the signal reaches the arm is slower. Hence, the disease may progress. In the second case, it indicates that the future latency is getting smaller than the normal range. It hints that the disease is getting better or going to the emitting phase. In any case, it suggests further tests are needed for those PwMS.

Figure 14 shows the IRI of 31 PwMS from the right foot data. It is observed that 22/25 (88%) PwMS with more than one future point can be detected with progression using IRI. There are 16/22 (72 %) that the future points are close or greater than the upper bound of the interval and 10/22 (45 %) that the future points are close or smaller than the lower bound of the interval. The same analogy applies to the result for the right hand data; these 22 PwMS need further tests or be cautious about the possibility of progression.

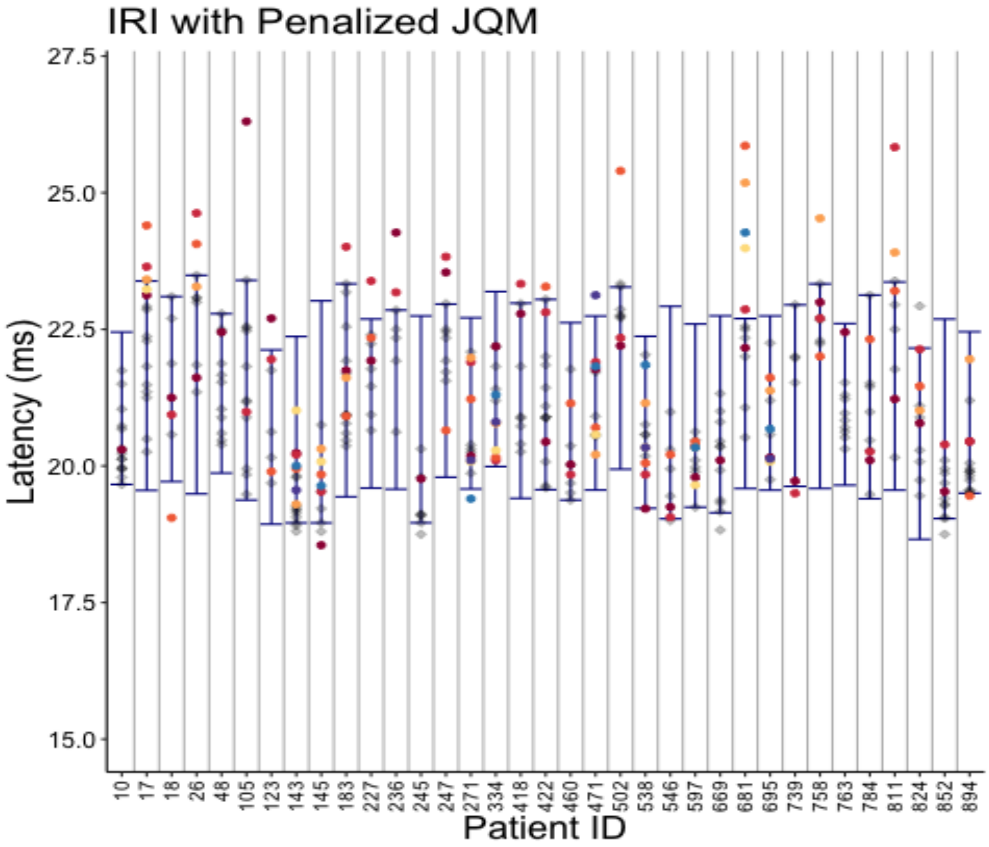


Figure 13: IRI for hand right data, with latency stability check. The blue intervals are the IRI. The gray diamonds are the stable data. The colored points are future data. The more blueish the point, the further in the future the data is.

In addition, in comparison to the results from the previous section (see Figure 11, 12, 19, 20), it is observed that the accuracy significantly decreases when there is no stability check. This confirms that the IRI is sensitive to the stability of the data, so data management needs to be done very carefully. Table 3 provides detailed information on the accuracy differences between the two cases, specifying the sample size in each subset.

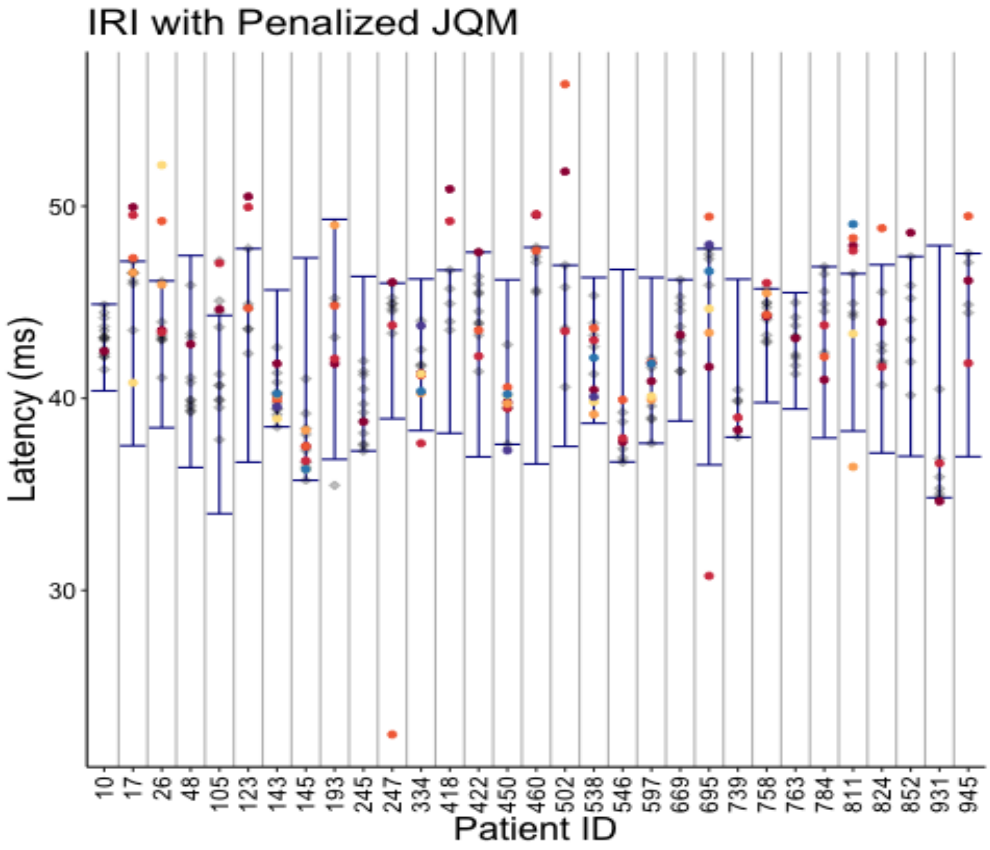


Figure 14: IRI for right foot data, with latency stability check. The blue intervals are the IRI. The gray diamonds are the stable data. The colored points are future data. The more blueish the point, the further in the future the data is.

Table 3: Accuracy of IRI with and without stability check with four subsets. The corresponding sample size is in brackets.

	Without stability check	**With stability check**
Right Hand	45% (48)	83% (34)
Right Foot	55% (43)	80% (31)
Left Hand	40% (48)	60% (29)
Left Foot	47% (43)	64% (33)

It is also observed that there are a few PwMS whose future points are totally inside the IRI. There are different reasons for this phenomenon. Firstly, it can be explained by the way *breakpoint* is defined. The *breakpoint* is calculated based on EDSS score. However,

the EDSS score and the latency are examined independently on different days, so there can be a gap between the time the last latency data is included in the stable data and the time when the first latency data is included in the future data. Another explanation is based on how the EDSS score is calculated. The EDSS score is based on foot mobility rather than hand mobility. Hence, there are cases of a patient with a high score of EDSS score, where the PwMS hardly walks, but the hand mobility is still quite good. Hence, the IRI using *breakpoint* based on EDSS score, will not capture 100% the progression of hands motility.

It has been observed that differences exist between the left and right sides of the body. As seen in Table 4, when covariates such as sex and age are not considered, higher accuracy is noted on the right side compared to the left side. However, the trending of the future points outside of IRI is more pronounced on the left side, meaning that the distance between the future points to the lower bound (upper bound) is much higher (refers to Figure 21 and Figure 22 on Appendix B).

4.3 IRI with stability check considering sex and age

Figure 15 shows the IRI of 34 PwMS from the right hand data. It is observed that 24/29 (83%) PwMS with more than one future point can be detected with progression using IRI. There are 14/24 (58 %) that the future points are close or greater than the upper bound of the interval and 10/24 (42 %) that the future points are close or smaller than the lower bound of the interval. The same analogy applies to this result: these 24 PwMS need further tests or be cautious about the possibility of progression.

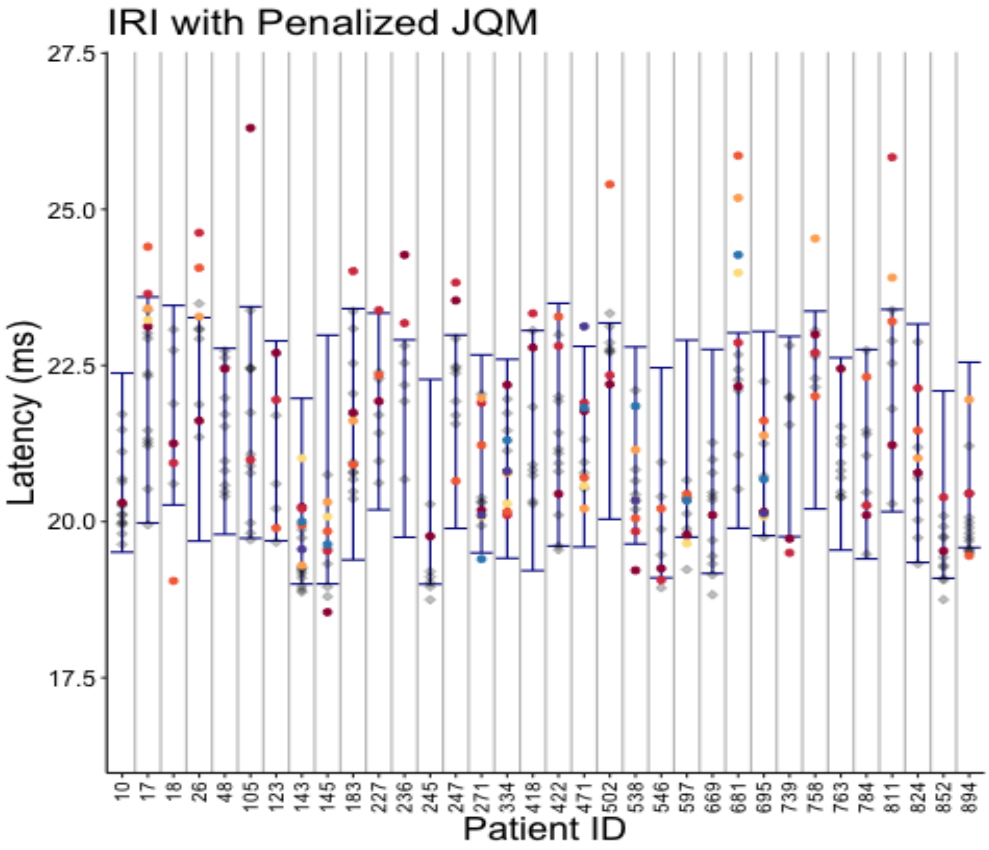


Figure 15: IRI for the right hand data, with stability check, considering sex and age covariates. The blue intervals are the IRI. The gray diamonds are the stable data. The colored points are future data. The more blueish the point, the further in the future the data is.

Figure 16 shows the IRI of 31 PwMS from the right foot data. It is observed that 16/25 (64%) PwMS with more than one future point can be detected with progression using IRI. There are 12/16 (75 %) that the future points are close or greater than the upper bound of the interval and 5/16 (31 %) that the future points are close or smaller than the lower bound of the interval. The same analogy applies to this result: these 16 PwMS need further tests or be cautious about the possibility of progression.

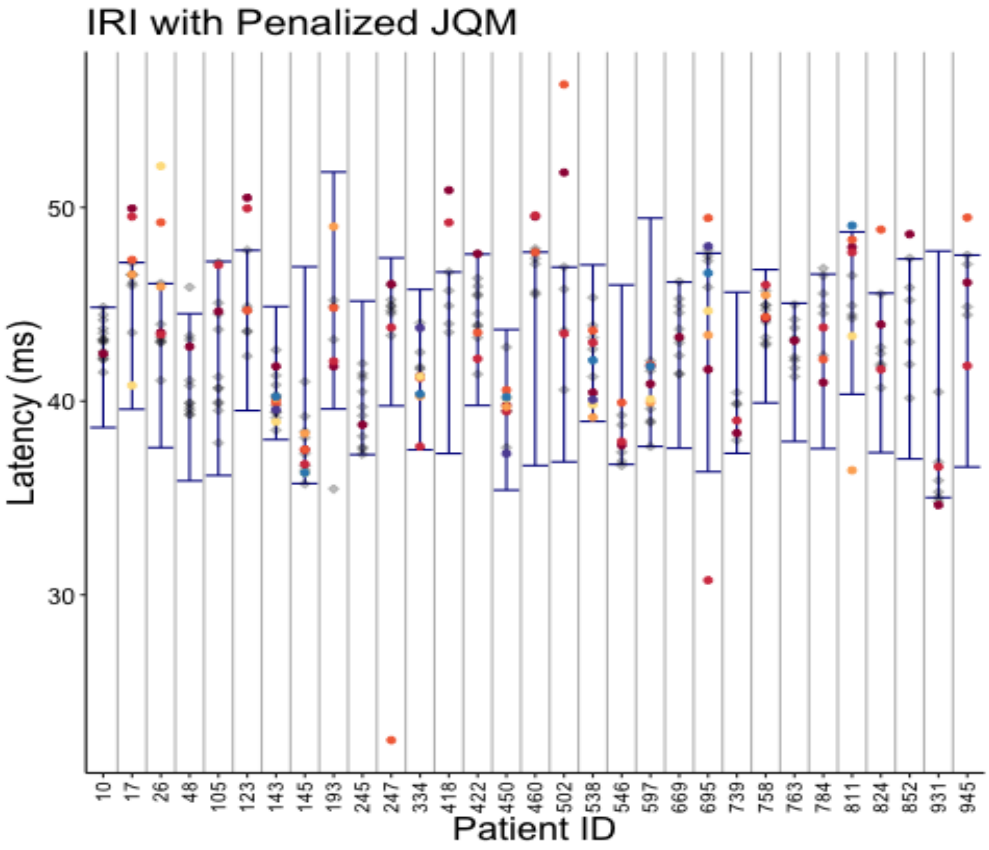


Figure 16: IRI for the right foot data, with stability check, considering sex and age covariates. The blue intervals are the IRI. The gray diamonds are the stable data. The colored points are future data. The more blueish the point, the further in the future the data is.

A higher accuracy on the right side, compared to the left side, was not observed when covariates were included in the calculation of IRI, as opposed to when covariates were not used. However, the deviation of future points outside of IRI is still more noticeable on the left side than on the right side, which is consistent with the case when no covariates were used (refer to Figure 23 and Figure 24 in Appendix C). Finally, in Table 4, the accuracy of IRI with and without including the covariate sex and age for four subdatasets is compared. Specifically, two out of four datasets show a noticeable improvement in accuracy, and one out of four datasets has the same accuracy. It hints that considering these factors can improve the predictive ability of IRI. However, it cannot be applied to the right hand dataset, where the opposite trend is observed. In short, caution should be exercised when interpreting the result of adding more covariates to improve accuracy, especially due to the small sample size in our case.

Table 4: Accuracy of IRI with and without covariates (sex and age) with four subsets. The corresponding sample size is in brackets. Sample size when considering covariates and when not are the same.

	Without covariate	**With covariate**
Right Hand	83% (34)	64%
Right Foot	80% (31)	88%
Left Hand	60% (29)	71%
Left Foot	64% (33)	64%

4.4 Confidence Interval of IRI

The IRI's 95% CI calculation is done using Bootstrap. As our sample size is relatively small, it is important to see how sensitive the estimated lower and upper boundaries of each PwMS change when the dataset changes. For each individual, it is observed that the lower and upper bounds are close to the median of their Bootstrap datasets. The distribution of the bounds for each patient is shown in detail in Appendix E and Appendix D. The preference for the median instead of the mean is due to the non-parametric nature of their distribution. Our prior results already confirm that IRI effectively detects progression in MS. Expanding to this foundation, it is demonstrated that compared to the standard EDSS score, using IRI on latency also provides significant advantages. Specifically, it is noticeable that in most PwMS, IRI can detect progression faster than EDSS score. Figure 17 and Figure 18 show six PwMS with the time to detect progression using EDSS score compared to the time using IRI. The findings for other PwMS can be found in Appendix I and Appendix H. The results for the left hand and the left foot are shown in Appendix J and Appendix K. It is worth mentioning that the EDSS score in Figure 17 and Figure 18 are transformed with a time interval of two years. Each individual's original EDSS score evolution and IRI can be found in Appendix L - Appendix O.

Moreover, it is shown in Table 5 and Table 6 the detail of detecting time using EDSS score versus using IRI. The calculation indicates a noticeable advantage in the early detection capacities of IRI. Specifically, on average, the IRI detects the progression 15.69 months earlier in the right hand dataset and 16.64 months in the right foot dataset. This significant time can be seen as an opportunity for timely intervention and adjustment of the management plan, potentially improving patient response.

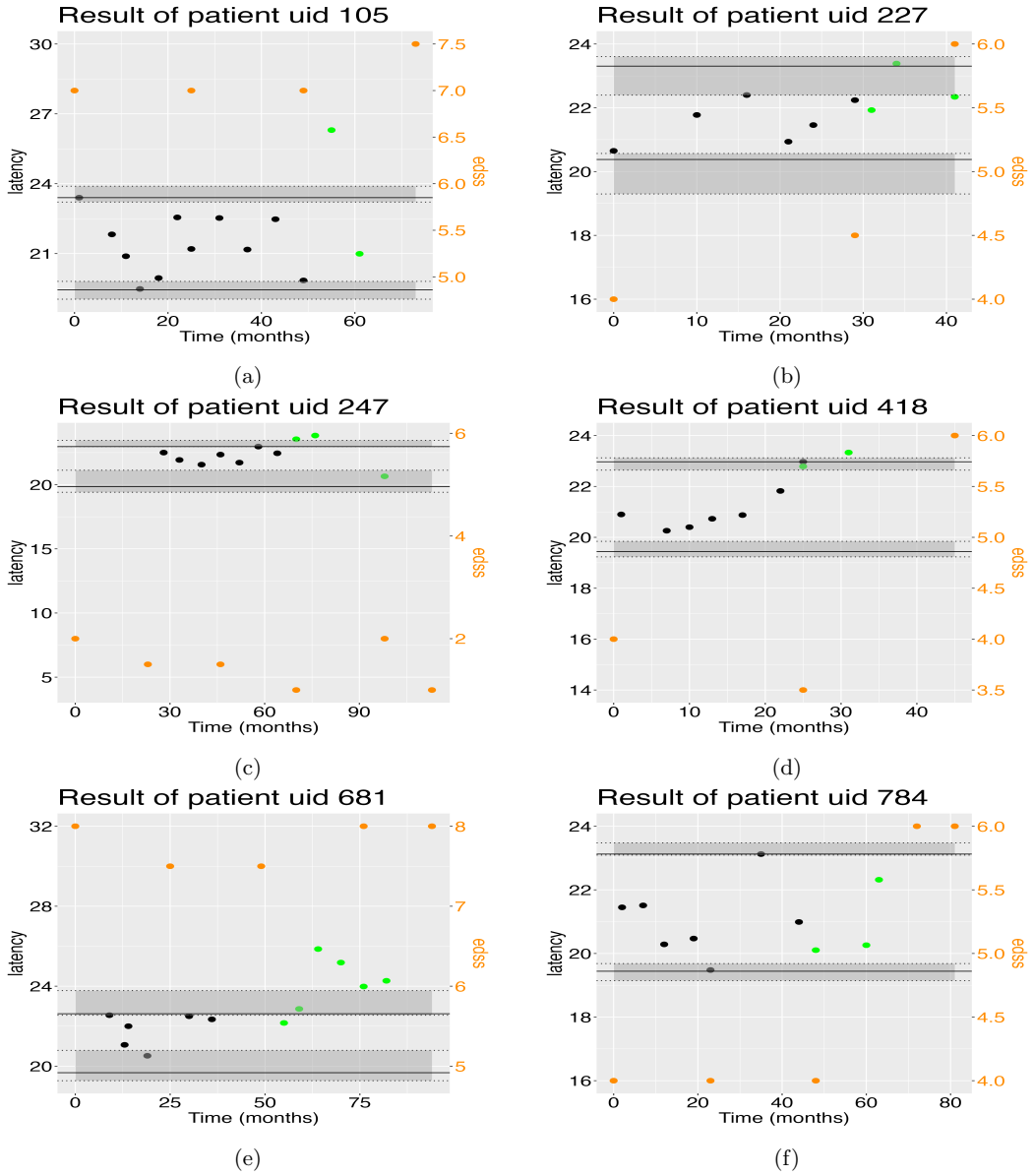


Figure 17: A subset of PwMS from the right foot data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression can be detected at 73 months based on EDSS score, and at 55 months based on IRI.
(b) Progression can be detected at 41 months based on EDSS score, and at 34 months based on IRI.
(c) Progression can be detected at 98 months based on EDSS score, and at 70 months based on IRI.
(d) Progression can be detected at 45 months based on EDSS score, and at 25 months based on IRI.
(e) Progression can be detected at 76 months based on EDSS score, and at 59 months based on IRI.
(f) Progression can be detected at 74 months based on EDSS score, and at 64 months based on IRI.

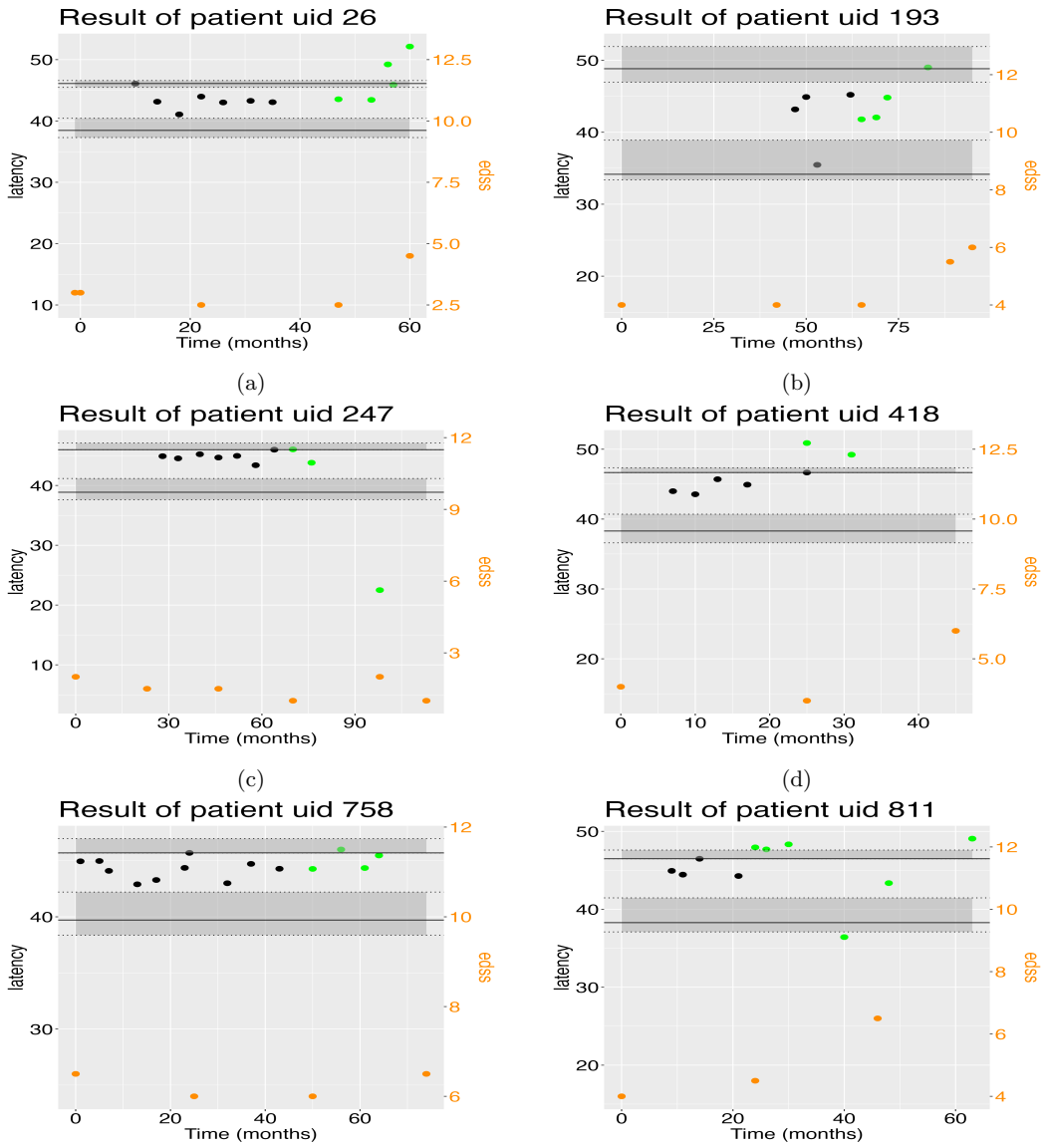


Figure 18: A subset of PwMS from the right foot data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression can be detected at 60 months based on EDSS score, and at 56 months based on IRI.
(b) Progression can be detected at 89 months based on EDSS score, and at 83 months based on IRI.
(c) Progression can be detected at 98 months based on EDSS score, and at 70 months based on IRI.
(d) Progression can be detected at 45 months based on EDSS score, and at 25 months based on IRI.
(e) Progression can be detected at 74 months based on EDSS score, and at 56 months based on IRI.
(f) Progression can be detected at 46 months based on EDSS score, and at 24 months based on IRI.

Table 5: Time (months) to detect progression using EDSS score and IRI from the right hand data. *Diff* column represents the difference between them. *No data available* is when there is no latency future data between the date of the last data point of stable data and the date to detect progression using EDSS score. *not able to detect progression* is when there is latency future data between the date of the last datapoint in stable data and the date to detect progression using EDSS score, but those data cannot detect progression. Stable data is all data collected before *breakpoint*. Future data is all data collected after *breakpoint*.

Patient_uid	**EDSS score**	**IRI**	**Diff**	**Patient_uid**	**EDSS score**	**IRI**	**Diff**
10	no data available			422	76	56	**20**
17	72	54	**18**	471	48	33	**15**
18	same time			502	99	89	**10**
26	60	56	**4**	538	49	38	**11**
48	no data available			546	50	28	**22**
105	73	55	**18**	597	not able to detect progression		
123	73	68	**5**	669	no data available		
143	74	56	**18**	681	76	59	**17**
145	not able to detect progression			695	50	32	**18**
183	122	102	**20**	739	39	23	**16**
227	41	34	**7**	758	74	64	**10**
236	no data available			763	57	37	**20**
245	no data available			784	not able to detect progression		
247	98	70	**28**	811	46	30	**16**
271	no data available			824	64	59	**5**
334	72	46	**26**	852	not able to detect progression		
418	45	25	**20**	894	92	75	**17**

Table 6: Time (months) to detect progression using EDSS score and IRI from the right foot data. *Diff* column represents the difference between them. *No data available* is when there is no latency future data between the date of the last data point of stable data and the date to detect progression using EDSS score. *not able to detect progression* is when there is latency future data between the date of the last datapoint in stable data and the date to detect progression using EDSS score, but those data cannot detect progression. Stable data is all data collected before *breakpoint*. Future data is all data collected after *breakpoint*.

Patient_uid	**EDSS score**	**IRI**	**Diff**	**Patient_uid**	**EDSS score**	**IRI**	**Diff**
10	not able to detect progression			502	99	77	**22**
17	72	50	**22**	538	49	29	**20**
26	60	56	**4**	546	not able to detect progression		
48	not able to detect progression			597	not able to detect progression		
105	73	55	**18**	669	no data available		
123	73	68	**15**	695	50	32	**18**
143	74	70	**4**	739	not able to detect progression		
145	73	68	**5**	758	74	56	**18**
193	89	83	**6**	763	not able to detect progression		
245	no data available			784	72	48	**24**
247	98	70	**28**	811	46	24	**22**
334	72	52	**20**	824	64	59	**5**
418	45	25	**20**	852	95	77	**18**
422	76	51	**25**	931	50	31	**19**
450	not able to detect progression			945	32	23	**9**
460	75	51	**24**				

5 Discussion

In this study, comprehensive data from PwMS were utilized to investigate the effectiveness of MEP measurement in detecting MS progression. A non-parametric method was performed to estimate the normal range of each individual using the longitudinal data. Evidence of high accuracy (60%-88%) was found using MEP. Moreover, compared to the traditional EDSS score, MEP detected progression on average 16 months earlier. The IRI was also identified as useful in detecting MS progression.

The accuracy of IRI using the hand and foot datasets shows a slight difference, as does the accuracy between the right and left sides, even when considering the inclusion or exclusion of age and sex as covariates. Despite this, the overarching conclusion regarding the effectiveness of utilizing IRI on MEP remains consistent.

Application in clinical context

It is clear that using IRI to determine the individual range of latency can help clinicians detect the progression of a patient's condition faster. Using this approach can also help clinicians to evaluate the current DMT in a better way. It is suggested to use it as a support tool, aiding in decision-making whether the ongoing DMT is effective for a patient or any change is needed, ensuring that PwMS can receive the optimal DMT in time. The application is available on: IRI Application.

Limitation and future research

The findings of this study have some limitations. First, *breakpoint* is defined based mostly on the first signal of disease progression. In this study, distinctions have not been made between RRMS, SPMS, and PMS. However, this information might be crucial in a clinical setting, particularly given that most PwMS are diagnosed with RRMS. In future research, taking into account this information, it is possible to more accurately and effectively evaluate the individual latency range for each patient, which may lead to different DMT adjustments depending on the MS type.

Secondly, it is assumed that the four subsets, comprising the left hand, right hand, left foot, and right foot, are independent. However, it is common sense that if the symptoms show in more than one subset from a patient, it may indicate a more complex situation. The fact that the subset sample sizes are not balanced and are small makes further study complicated.

In terms of prediction, all "abnormal" data are excluded, which may exclude early signs of progression, based on the MEP measurements. On the other hand, the running time for the IRI is significant, especially when adding more covariates. Hence, a good-performance

computer may be needed when the sample size increases.

In our study, it is decided not to implement the Bootstrap method for the IRI, which includes the covariates of sex and age. This decision was based on the computation time estimation would be excessively long, thus making the process impractical for our timeline. Nevertheless, our preliminary observations indicated a difference in accuracy dependent on whether these covariates were included or excluded. This suggests that incorporating these factors could potentially give more insights into our findings. Therefore, despite the challenges, it may be valuable to explore this method in future research to fully understand the impact of these covariates on the IRI.

In addition, it is important to note that IRI is sensitive to data stability. When the threshold of the latency measurement is loosened, the accuracy decreases significantly (from 17% to 38%). As shown in the results section, it's crucial to perform the IRI with a sufficiently large sample size to ensure reliability. Although the algorithm can work with a minimum of three data points from each patient, it is recommended to have a sample size of at least 35 PwMS and four data points for each PwMS to achieve good accuracy. This larger sample size compensates for variability and ensures the IRI results are robust and reliable.

This study primarily examines latency measurement, but altitude measurement is also promising information for assessing the progression of MS. However, altitude measurement contains substantial variation, even within individuals. Therefore, in order to pursue this, the study's design would need to be adjusted. Unfortunately, this is not feasible within the limited time available.

Finally, this study focuses on MEP using IRI as a supported instrument next to EDSS. To extend the result to use IRI on MEP as an independent test, one need to calculate the specificity, not only the sensitivity as shown above.

References

Multiple Sclerosis Diagnosis and Treatment. MDPI - Multidisciplinary Digital Publishing Institute, Basel, 2022. ISBN 978-3-0365-3097-0. OCLC: 1302716397.

S. L. P. Apóstolos, M. Boaventura, N. T. Mendes, L. S. Teixeira, and I. G. Campana. How to choose initial treatment in multiple sclerosis patients: a case-based approach. *Arquivos de Neuro-Psiquiatria*, 80(5 suppl 1):159–172, May 2022. ISSN 1678-4227, 0004-282X. doi: 10.1590/0004-282x-anp-2022-s128. URL http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0004-282X2022000700159&tlng=en.

o. M. Atlas. Number of people with MS | Atlas of MS. URL https://www.atlasofms.org/map/global/epidemiology/number-of-people-with-ms.

G. R. Cutter. Development of a multiple sclerosis functional composite as a clinical trial outcome measure. *Brain*, 122(5):871–882, May 1999. ISSN 14602156. doi: 10.1093/brain/122.5.871. URL https://academic.oup.com/brain/article-lookup/doi/10.1093/brain/122.5.871.

E. De Brouwer, T. Becker, Y. Moreau, E. K. Havrdova, M. Trojano, S. Eichau, S. Ozakbas, M. Onofrj, P. Grammond, J. Kuhle, L. Kappos, P. Sola, E. Cartechini, J. Lechner-Scott, R. Alroughani, O. Gerlach, T. Kalincik, F. Granella, F. Grand'Maison, R. Bergamaschi, M. José Sá, B. Van Wijmeersch, A. Soysal, J. L. Sanchez-Menoyo, C. Solaro, C. Boz, G. Iuliano, K. Buzzard, E. Aguera-Morales, M. Terzi, T. C. Trivio, D. Spitaleri, V. Van Pesch, V. Shaygannejad, F. Moore, C. Oreja-Guevara, D. Maimone, R. Gouider, T. Csepany, C. Ramo-Tello, and L. Peeters. Longitudinal machine learning modeling of MS patient trajectories improves predictions of disability progression. *Computer Methods and Programs in Biomedicine*, 208:106180, Sept. 2021. ISSN 01692607. doi: 10.1016/j.cmpb.2021.106180. URL https://linkinghub.elsevier.com/retrieve/pii/S0169260721002546.

V. Fernández, J. Valls-Sole, J. Relova, N. Raguer, F. Miralles, L. Dinca, S. Taramundi, L. Costa-Frossard, M. Ferrandiz, L. Ramió-Torrentà, P. Villoslada, A. Saiz, C. Calles, A. Antigüedad, J. Alvarez-Cermeño, J. Prieto, G. Izquierdo, X. Montalbán, and O. Fernández. Recommendations for the clinical use of motor evoked potentials in multiple sclerosis. *Neurología (English Edition)*, 28(7):408–416, Sept. 2013. ISSN 21735808. doi: 10.1016/j.nrleng.2013.09.003. URL https://linkinghub.elsevier.com/retrieve/pii/S2173580813001247.

J. S. Fischer, R. A. Rudick, G. R. Cutter, and S. C. Reingold. The Multiple Sclerosis Functional Composite measure (MSFC): an integrated approach to MS clinical outcome assessment.

M. S. Freedman, V. Devonshire, P. Duquette, P. S. Giacomini, F. Giuliani, M. C. Levin, X. Montalban, S. A. Morrow, J. Oh, D. Rotstein, and E. A. Yeh. Treatment Optimization in Multiple Sclerosis: Canadian MS Working Group Recommendations. *Canadian Journal of Neurological Sciences / Journal Canadien des Sciences Neurologiques*, 47(4): 437–455, July 2020. ISSN 0317-1671, 2057-0155. doi: 10.1017/cjn.2020.66. URL https://www.cambridge.org/core/product/identifier/S0317167120000669/type/journal_article.

G. Giovannoni and R. W. Rhoades. Individualizing treatment goals and interventions for people with MS. *Current Opinion in Neurology*, 25:S20–S27, Feb. 2012. ISSN 1350-7540. doi: 10.1097/01.wco.0000413321.32834.aa. URL https://journals.lww.com/00019052-201202001-00004.

R. Gold, J. S. Wolinsky, M. P. Amato, and G. Comi. Evolving expectations around early management of multiple sclerosis. *Therapeutic Advances in Neurological Disorders*, 3(6): 351–367, Nov. 2010. ISSN 1756-2864, 1756-2864. doi: 10.1177/1756285610385608. URL http://journals.sagepub.com/doi/10.1177/1756285610385608.

M. Haki, H. A. AL-Biati, Z. S. Al-Tameemi, I. S. Ali, and H. A. Al-hussaniy. Review of multiple sclerosis: Epidemiology, etiology, pathophysiology, and treatment. *Medicine*, 103 (8):e37297, Feb. 2024. ISSN 0025-7974, 1536-5964. doi: 10.1097/MD.0000000000037297. URL https://journals.lww.com/10.1097/MD.0000000000037297.

M. Hardmeier, L. Leocani, and P. Fuhr. A new role for evoked potentials in MS? Repurposing evoked potentials as biomarkers for clinical trials in MS. *Multiple Sclerosis Journal*, 23 (10):1309–1319, Sept. 2017. ISSN 1352-4585, 1477-0970. doi: 10.1177/1352458517707265. URL http://journals.sagepub.com/doi/10.1177/1352458517707265.

M. Hardmeier, C. Schindler, J. Kuhle, and P. Fuhr. Validation of Quantitative Scores Derived From Motor Evoked Potentials in the Assessment of Primary Progressive Multiple Sclerosis: A Longitudinal Study. *Frontiers in Neurology*, 11:735, July 2020. ISSN 1664-2295. doi: 10.3389/fneur.2020.00735. URL https://www.frontiersin.org/article/10.3389/fneur.2020.00735/full.

S. I. Hauser, D. M. Dawson, and J. R. Leurich. Intensive Immunosuppression in Progressive Multiple Sclerosis. *The New ENgland Journal of Medicine*, 308(4):173–180, Jan. 1983. doi: 10.1056/NEJM198301273080401.

M. Hohol, E. Orav, and H. Weiner. Disease steps in multiple sclerosis: a longitudinal study comparing Disease Steps and EDSS to evaluate disease progression. 1999.

J. Hutchinson and M. Hutchinson. The functional limitations profile may be a valid, reliable and sensitive measure of disability in multiple sclerosis. *Journal of Neurology*, 242(10): 650–657, 1995. ISSN 0340-5354, 1432-1459. doi: 10.1007/BF00866915. URL http://link.springer.com/10.1007/BF00866915.

V. H. J. Geëvokeerde potentialen, 2007.

J. J. Kragt, J. M. Nielsen, F. A. Van Der Linden, B. M. Uitdehaag, and C. H. Polman. How similar are commonly combined criteria for EDSS progression in multiple sclerosis? *Multiple Sclerosis Journal*, 12(6):782–786, Nov. 2006. ISSN 1352-4585, 1477-0970. doi: 10.1177/1352458506070931. URL http://journals.sagepub.com/doi/10.1177/1352458506070931.

J. F. Kurtzke. Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology*, 33(11):1444–1444, Nov. 1983. ISSN 0028-3878, 1526-632X. doi: 10.1212/WNL.33.11.1444. URL https://www.neurology.org/doi/10.1212/WNL.33.11.1444.

L. Leocani, S. Medaglini, and G. Comi. Evoked potentials in monitoring multiple sclerosis. *Neurological Sciences*, 21(0):S889–S891, Dec. 2000. ISSN 1590-1874, 1590-3478. doi: 10.1007/s100720070032. URL http://link.springer.com/10.1007/s100720070032.

F. London, S. El Sankari, and V. Van Pesch. Early disturbances in multimodal evoked potentials as a prognostic factor for long-term disability in relapsing-remitting multiple sclerosis patients. *Clinical Neurophysiology*, 128(4):561–569, Apr. 2017. ISSN 13882457. doi: 10.1016/j.clinph.2016.12.029. URL https://linkinghub.elsevier.com/retrieve/pii/S1388245717300196.

J. J. Marriott, O. Ekuma, R. Fransoo, and R. A. Marrie. Switching to second line MS disease-modifying therapies is associated with decreased relapse rate. *Frontiers in Neurology*, 14:1243589, Sept. 2023. ISSN 1664-2295. doi: 10.3389/fneur.2023.1243589. URL https://www.frontiersin.org/articles/10.3389/fneur.2023.1243589/full.

S. Meyer-Moock, Y.-S. Feng, M. Maeurer, F.-W. Dippel, and T. Kohlmann. Systematic literature review and validity evaluation of the Expanded Disability Status Scale (EDSS) and the Multiple Sclerosis Functional Composite (MSFC) in patients with multiple sclerosis. 2014.

R. Milo and A. Miller. Revised diagnostic criteria of multiple sclerosis. *Autoimmunity Reviews*, 13(4-5):518–524, Apr. 2014. ISSN 15689972. doi: 10.1016/j.autrev.2014.01.012. URL https://linkinghub.elsevier.com/retrieve/pii/S156899721400024X.

J. H. Noseworthy, M. K. Vandervoort, C. J. Wong, and G. C. Ebers. Interrater variability with the Expanded Disability Status Scale (EDSS) and Functional Systems (FS) in a multiple sclerosis clinical trial. *Neurology*, 40(6):971–971, June 1990. ISSN 0028-3878, 1526-632X. doi: 10.1212/WNL.40.6.971. URL https://www.neurology.org/doi/10.1212/WNL.40.6.971.

C. H. Polman, S. C. Reingold, B. Banwell, M. Clanet, J. A. Cohen, M. Filippi, K. Fujihara, E. Havrdova, M. Hutchinson, L. Kappos, F. D. Lublin, X. Montalban, P. O'Connor, M. Sandberg-Wollheim, A. J. Thompson, E. Waubant, B. Weinshenker, and J. S. Wolinsky. Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria. *Annals of Neurology*, 69(2):292–302, Feb. 2011. ISSN 0364-5134, 1531-8249. doi: 10.1002/ana.22366. URL https://onlinelibrary.wiley.com/doi/10.1002/ana.22366.

M. Pusparum, G. Ertaylan, and O. Thas. From Population to Subject-Specific Reference Intervals. In V. V. Krzhizhanovskaya, G. Závodszky, M. H. Lees, J. J. Dongarra, P. M. A. Sloot, S. Brissos, and J. Teixeira, editors, *Computational Science – ICCS 2020*, volume 12140, pages 468–482. Springer International Publishing, Cham, 2020. ISBN 978-3-030-50422-9 978-3-030-50423-6. doi: 10.1007/978-3-030-50423-6_35. URL https://link.springer.com/10.1007/978-3-030-50423-6_35. Series Title: Lecture Notes in Computer Science.

M. Pusparum, G. Ertaylan, and O. Thas. Individual Reference Intervals for Personalized Interpretation of Clinical and Metabolomics Measurements. preprint, Health Informatics, June 2021. URL http://medrxiv.org/lookup/doi/10.1101/2021.06.04.21258334.

M. Ravnborgl, M. Grønbech-Jensen, and A. Jønsson. The MS impairment scale: a pragmatic approach to the assessment of impairment in patients with multiple sclerosis. *Multiple Sclerosis Journal*, 3(1):31–41, Feb. 1997. ISSN 1352-4585, 1477-0970. doi: 10.1177/135245859700300104. URL http://journals.sagepub.com/doi/10.1177/135245859700300104.

D. Robertson and N. Moreo. Disease-Modifying Therapies in Multiple Sclerosis: Overview and Treatment Considerations.

B. Sharrack and R. A. Hughes. The Guy's Neurological Disability Scale (GNDS): a new disability measure for multiple sclerosis.

J. C. Sipe, R. L. Knobler, S. L. Braheny, G. P. Rice, H. S. Panitch, and M. B. Oldstone. A neurologic rating scale (NRS) for use in multiple sclerosis. *Neurology*, 34(10):1368–1368, Oct. 1984. ISSN 0028-3878, 1526-632X. doi: 10.1212/WNL.34.10.1368. URL https://www.neurology.org/doi/10.1212/WNL.34.10.1368.

D. Stefanicka-Wojtas and D. Kurpas. Personalised Medicine—Implementation to the Healthcare System in Europe (Focus Group Discussions). *Journal of Personalized Medicine*, 13(3):380, Feb. 2023. ISSN 2075-4426. doi: 10.3390/jpm13030380. URL https://www.mdpi.com/2075-4426/13/3/380.

WHO. Multiple sclerosis. URL https://www.who.int/news-room/fact-sheets/detail/multiple-sclerosis.

J. Yperman, T. Becker, D. Valkenborg, N. Hellings, M. Cambron, D. Dive, G. Laureys, V. Popescu, B. Van Wijmeersch, and L. M. Peeters. Deciphering the Morphology of Motor Evoked Potentials. *Frontiers in Neuroinformatics*, 14:28, July 2020a. ISSN 1662-5196. doi: 10.3389/fninf.2020.00028. URL https://www.frontiersin.org/article/10.3389/fninf.2020.00028/full.

J. Yperman, T. Becker, D. Valkenborg, V. Popescu, N. Hellings, B. V. Wijmeersch, and L. M. Peeters. Machine learning analysis of motor evoked potential time series to predict disability progression in multiple sclerosis. *BMC Neurology*, 20(1):105, Dec. 2020b. ISSN 1471-2377. doi: 10.1186/s12883-020-01672-w. URL https://bmcneurol.biomedcentral.com/articles/10.1186/s12883-020-01672-w.

J. Yperman, V. Popescu, B. Van Wijmeersch, T. Becker, and L. M. Peeters. Motor evoked potentials for multiple sclerosis, a multiyear follow-up dataset. *Scientific Data*, 9(1): 207, May 2022. ISSN 2052-4463. doi: 10.1038/s41597-022-01335-0. URL https://www.nature.com/articles/s41597-022-01335-0.

Appendices

A IRI without latency stability

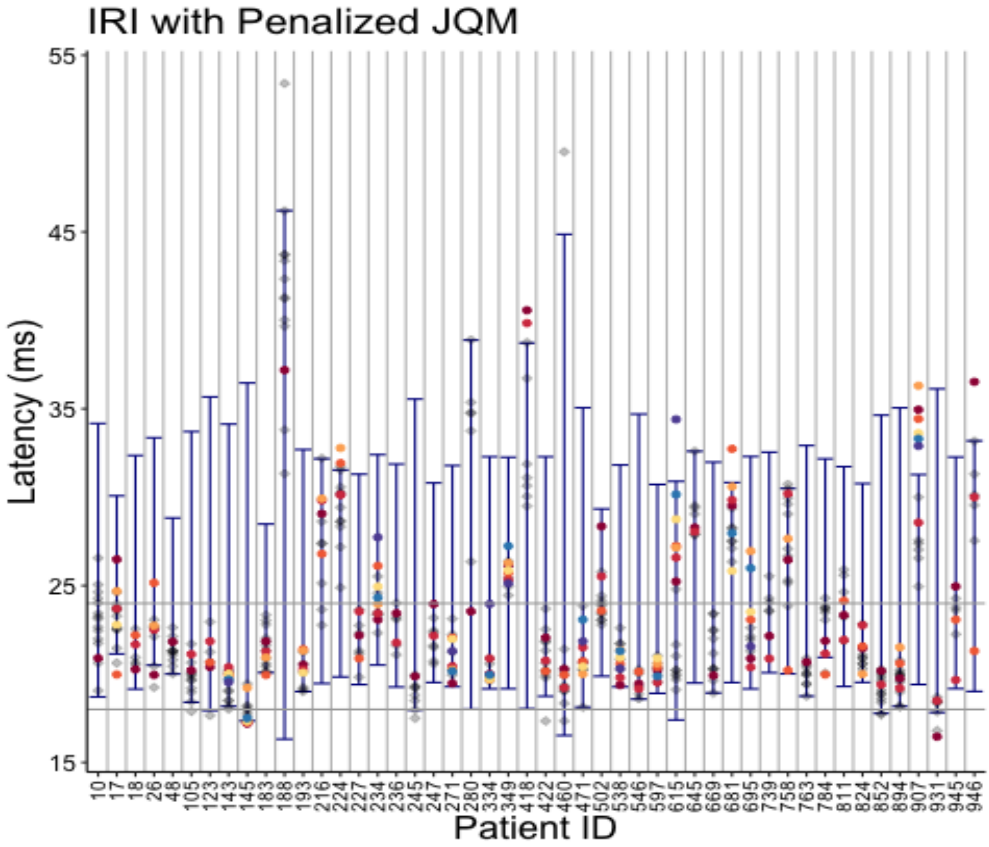


Figure 19: IRI from the left hand data, without stability for latency check. The blue intervals are the IRI. The gray diamonds are the stable data. The colored points are future data. The more blueish the point, the further in the future the data is. The two horizontal gray lines represent the population reference interval.

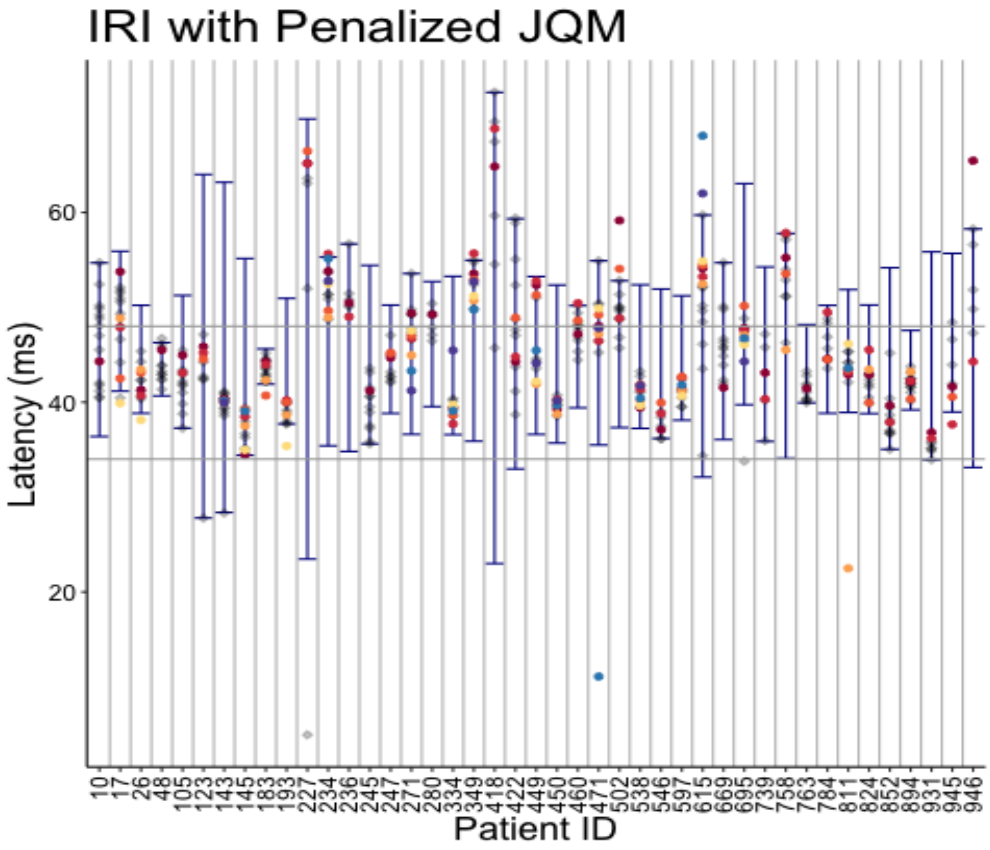


Figure 20: IRI from the left foot data, without stability for latency check. The blue intervals are the IRI. The gray diamonds are the stable data. The colored points are future data. The more blueish the point, the further in the future the data is. The two horizontal gray lines represent the population reference interval.

Figure 19 and Figure 20 show the IRI when there is no control on latency stability. IRI detects the progression when there is at least one future value (colored round points) that is close to the extremes or outside of the interval. For the left hand data, 48 PwMS are included, where 18/44 PwMS (40%) with more than one future point are detected with progression using IRI. For the left foot data, 43 PwMS are included, where 18/38 PwMS (47%) are detected with progression using IRI.

B IRI with latency stability

Figure 21 shows the IRI of 29 PwMS from the left hand data. It is observed that 15/25 (60%) PwMS with more than one future point can be detected with progression using IRI. There are 10/15 (66 %) that the future points are close or greater than the upper bound of the interval and 7/15 (46 %) that the future points are close or smaller than the lower bound of the interval. The same analogy applies to this result: these 16 PwMS need further tests or be cautious about the possibility of progression.

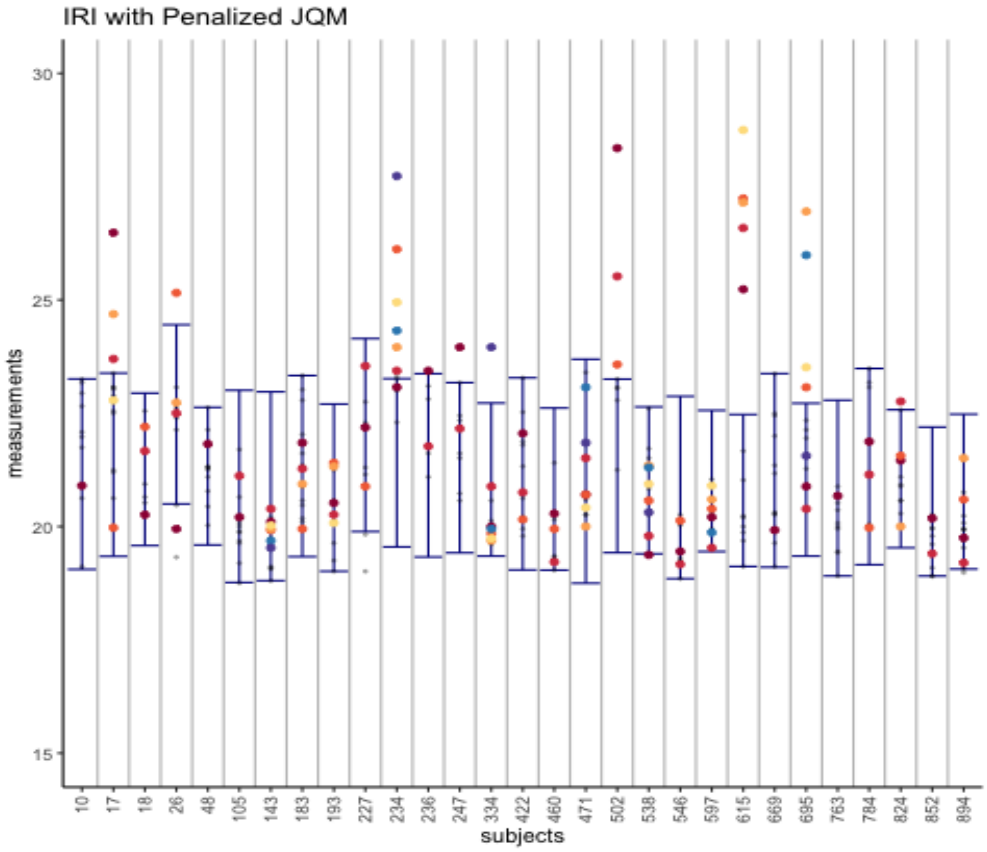


Figure 21: IRI for the left hand data with latency stability. The blue intervals are the IRI. The gray diamonds are the stable data. The colored points are future data. The more blueish the point, the further in the future the data is.

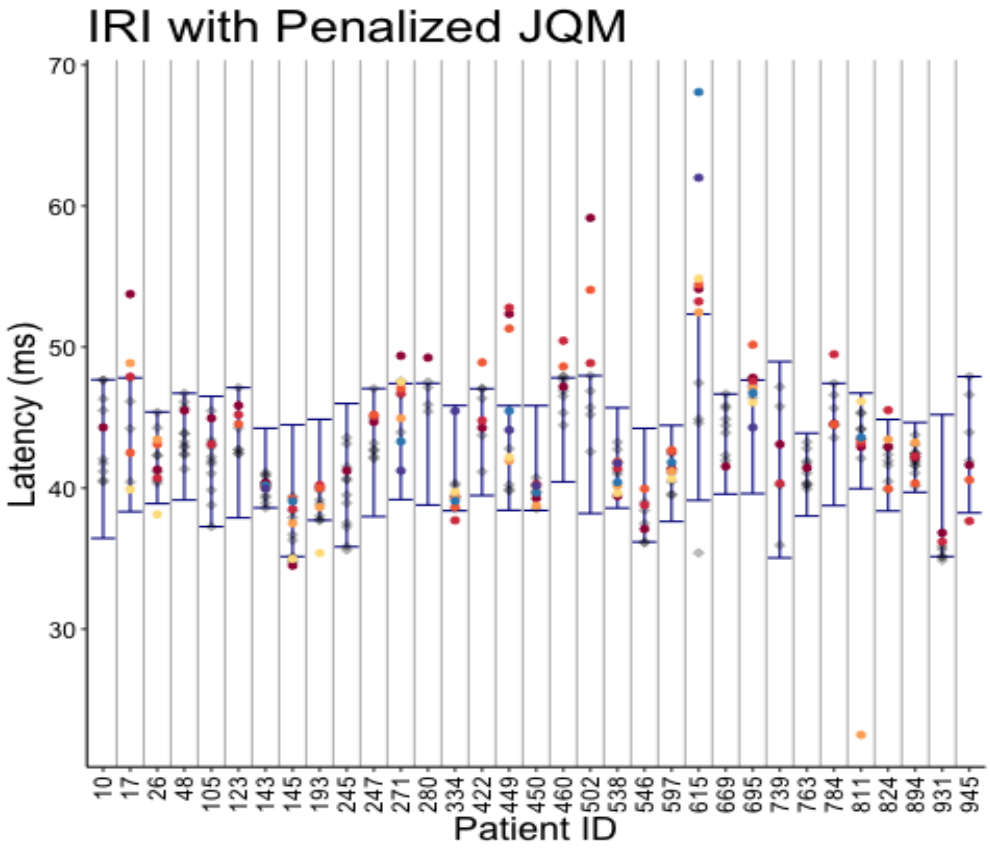


Figure 22: IRI for the left foot data with latency stability. The blue intervals are the IRI. The gray diamonds are the stable data. The colored points are future data. The more blueish the point, the further in the future the data is.

Figure 22 shows the IRI of 33 PwMS from the left foot data. It is observed that 18/28 (64%) PwMS with more than one future point can be detected with progression using IRI. There are 13/18 (72 %) that the future points are close or greater than the upper bound of the interval and 6/18 (33 %) that the future points are close or smaller than the lower bound of the interval. The same analogy applies to this result: these 18 PwMS need further tests or be cautious about the possibility of progression.

C IRI with covariates

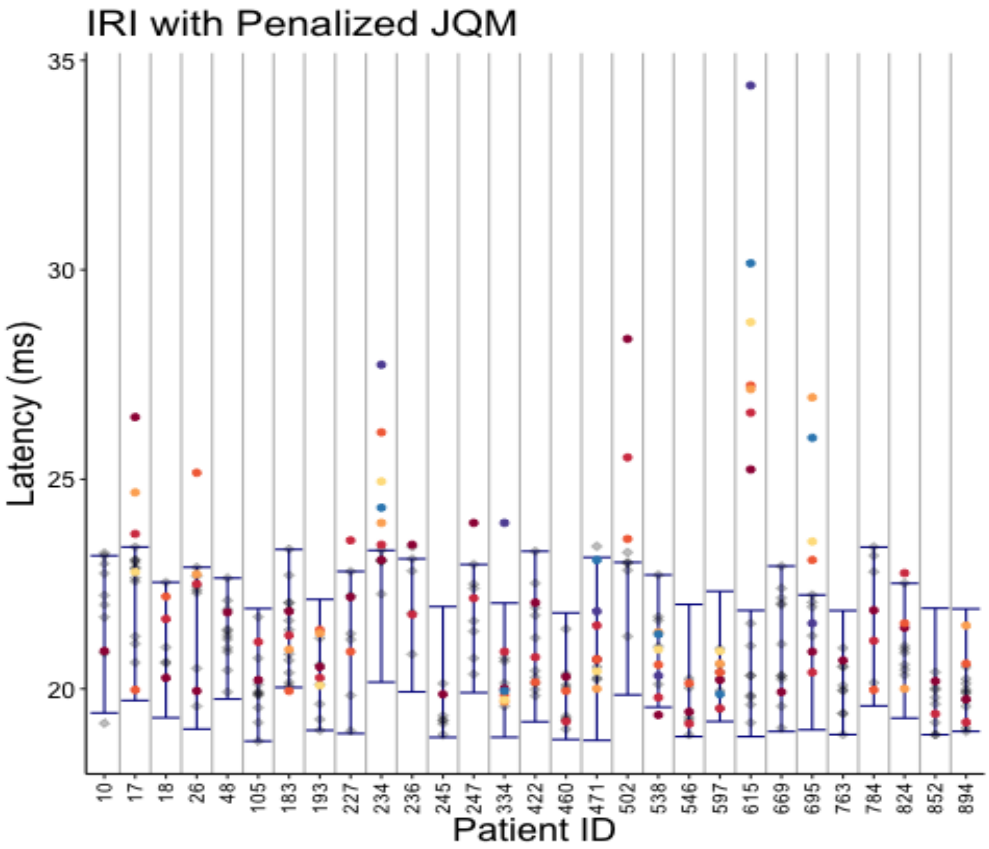


Figure 23: IRI for the left hand data, with stability check, considering sex and age covariates. The blue intervals are the IRI. The gray diamonds are the stable data. The colored points are future data. The more blueish the point, the further in the future the data is.

Figure 23 shows the IRI of 29 PwMS from the left hand data. It is observed that 17/24 (71%) PwMS with more than one future point can be detected with progression using IRI. There are 12/17 (70 %) that the future points are close or greater than the upper bound of the interval and 5/17 (30 %) that the future points are close or smaller than the lower bound of the interval. The same analogy applies to this result: these 17 PwMS need further tests or be cautious about the possibility of progression.

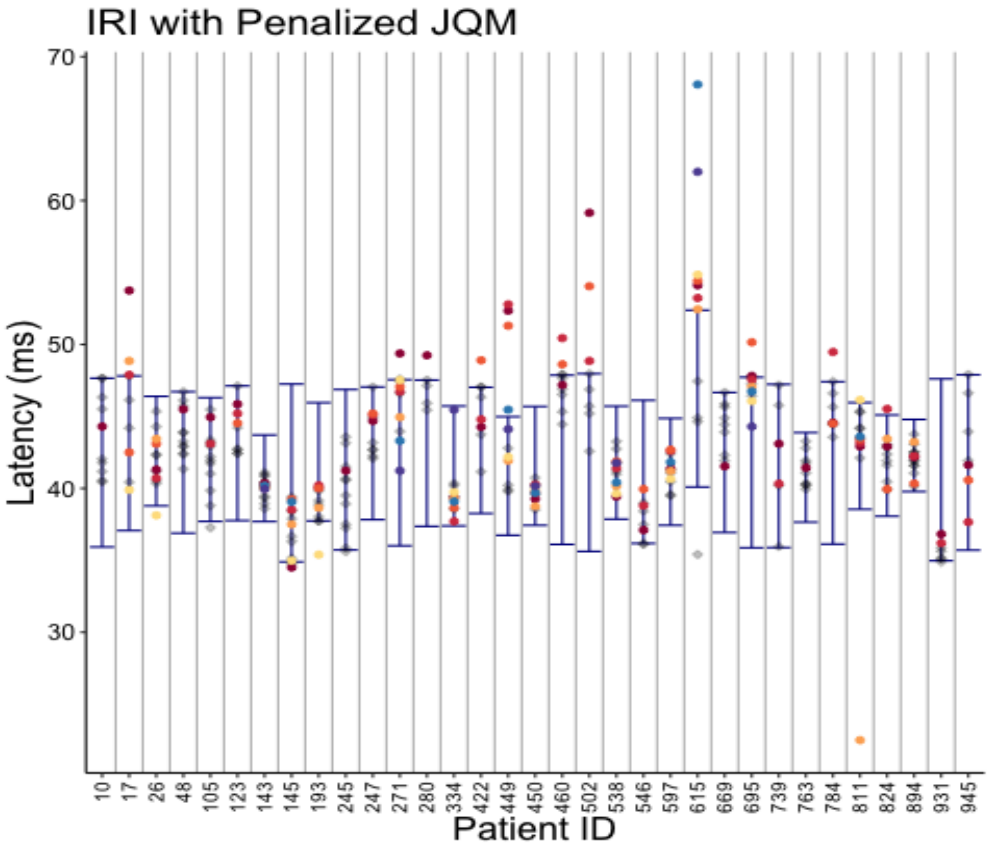


Figure 24: IRI for the left foot data, with stability check, considering sex and age covariates. The blue intervals are the IRI. The gray diamonds are the stable data. The colored points are future data. The more blueish the point, the further in the future the data is.

Figure 24 shows the IRI of 31 PwMS from the right foot data. It is observed that 16/25 (64%) PwMS with more than one future point can be detected with progression using IRI. There are 12/16 (75 %) that the future points are close or greater than the upper bound of the interval and 5/16 (31 %) that the future points are close or smaller than the lower bound of the interval. The same analogy applies to this result: these 16 PwMS need further tests or be cautious about the possibility of progression.

D Confidence Interval of IRI for right hand data

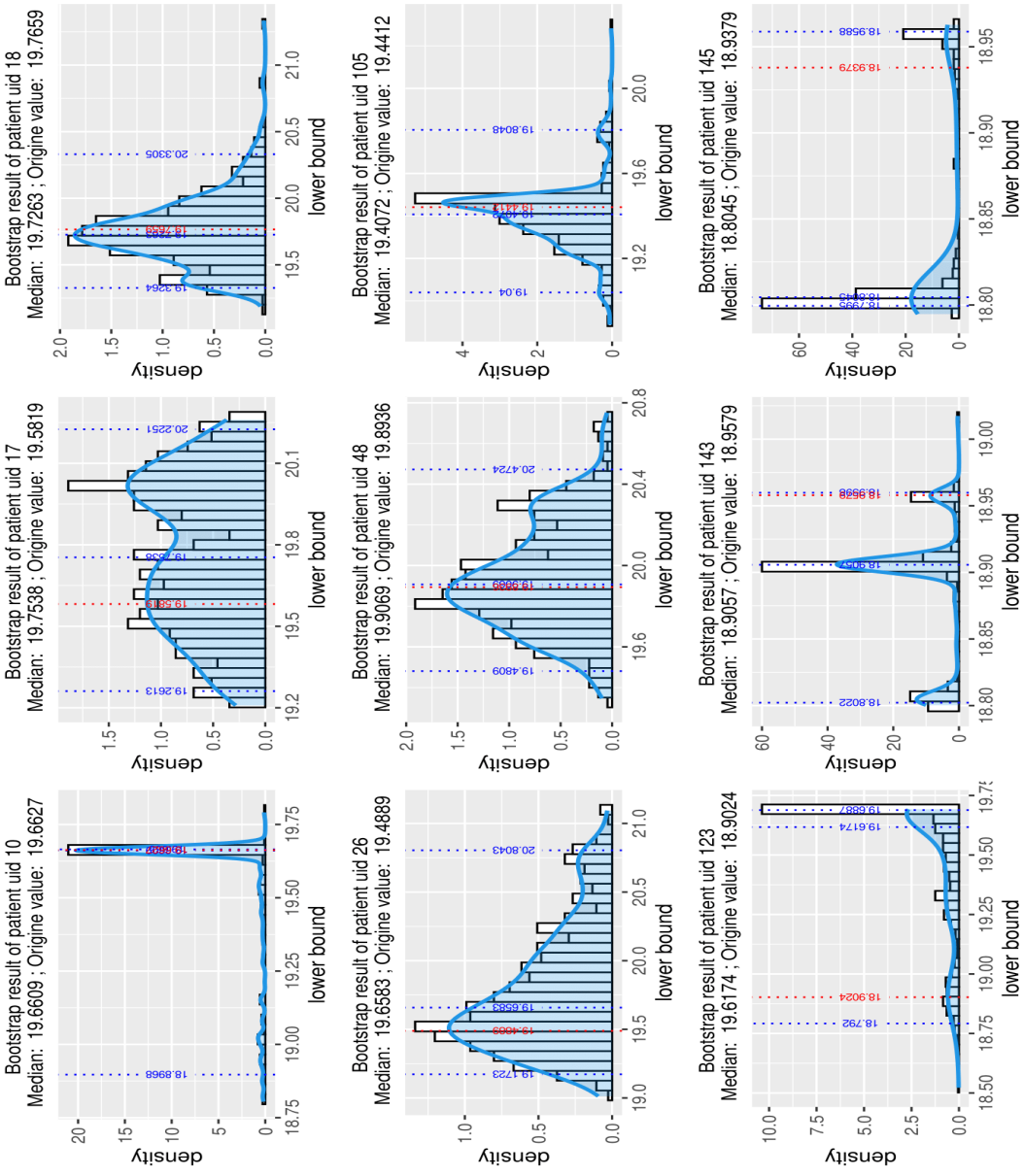


Figure 25: Lower bound page 1 for the right hand data. Histogram of the lower bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the origin value and the median are quite close.

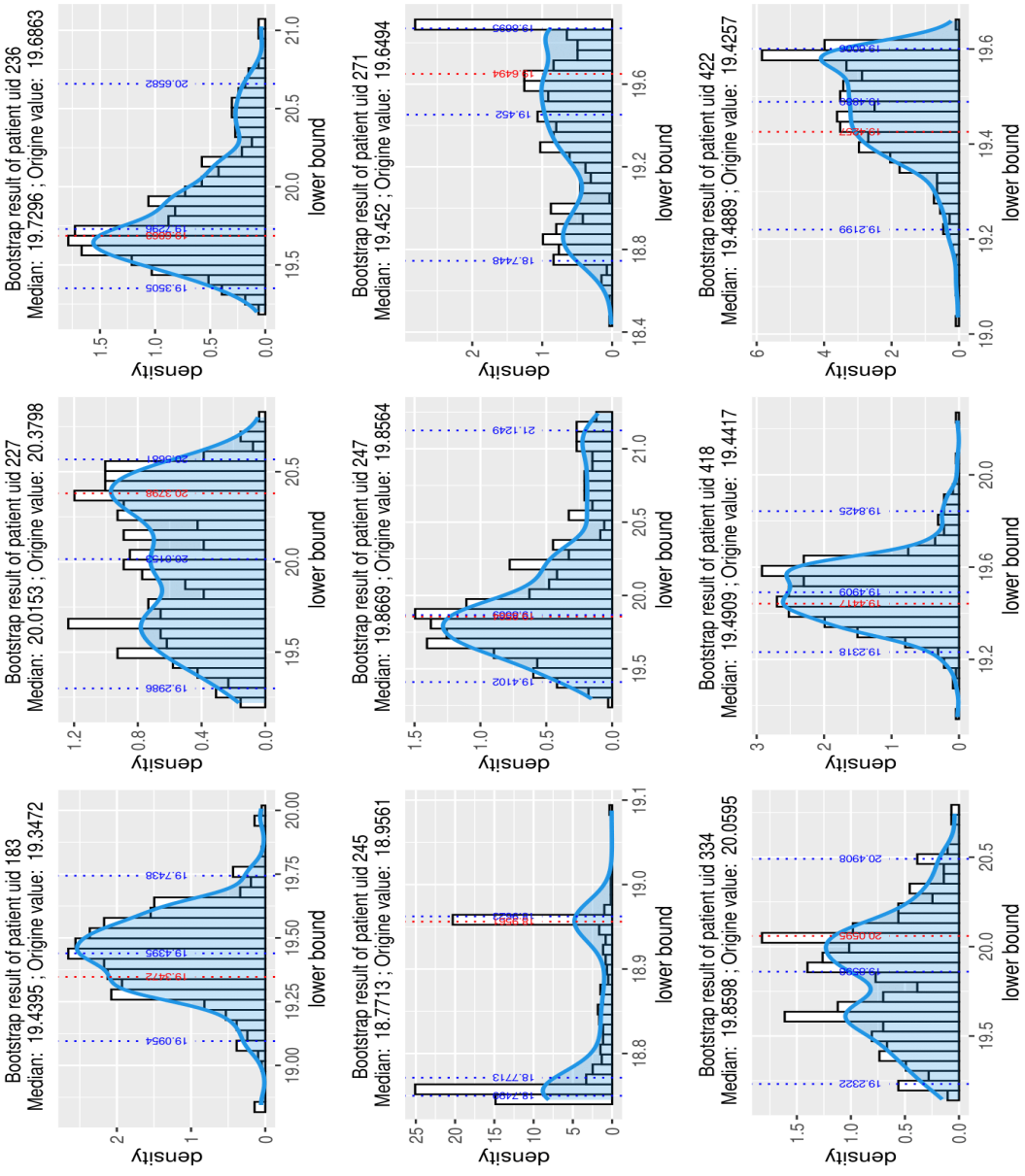


Figure 26: Lower bound page 2 for the right hand data. Histogram of the lower bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

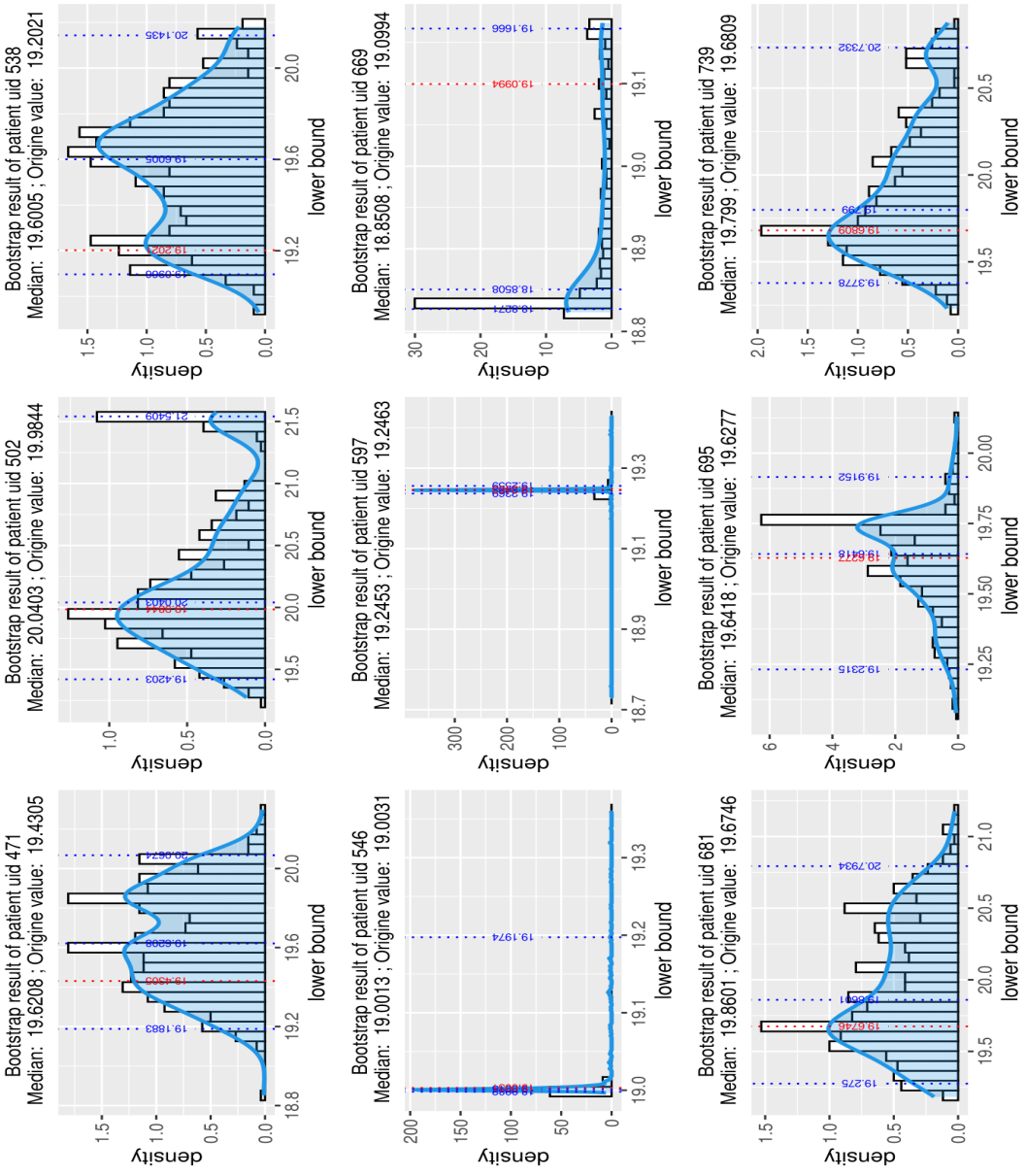


Figure 27: Lower bound page 3 for the right hand data. Histogram of the lower bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

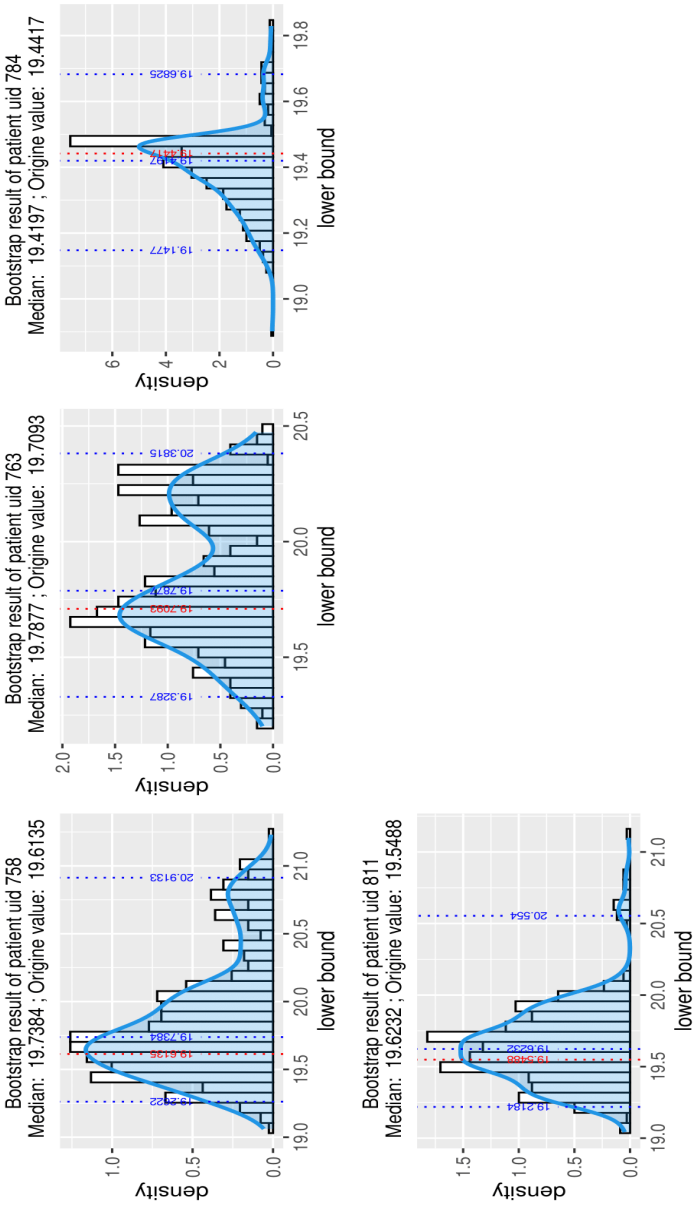


Figure 28: Lower bound page 4 for the right hand data. Histogram of the lower bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

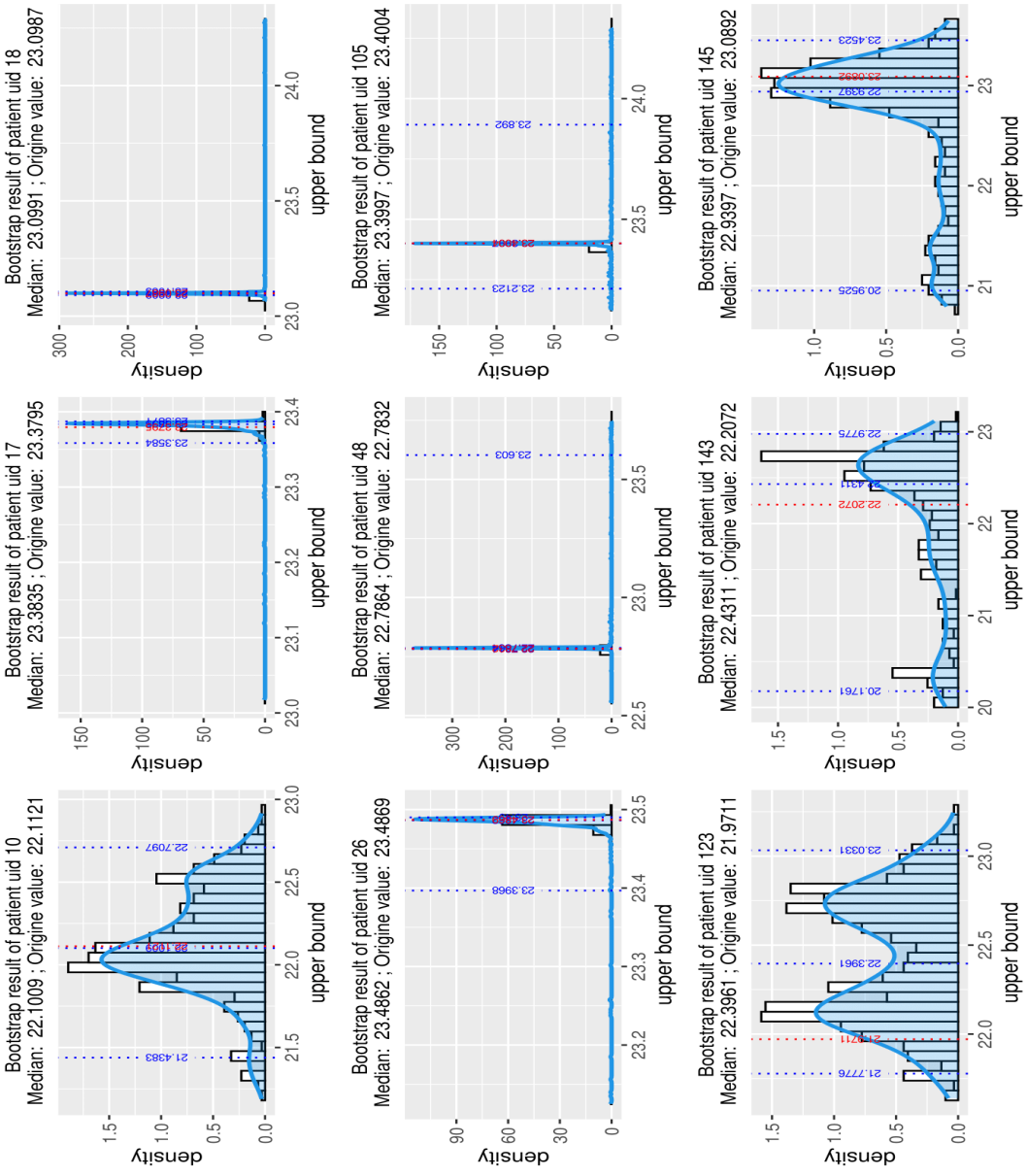


Figure 29: Upper bound page 1 for the right hand data. Histogram of the upper bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

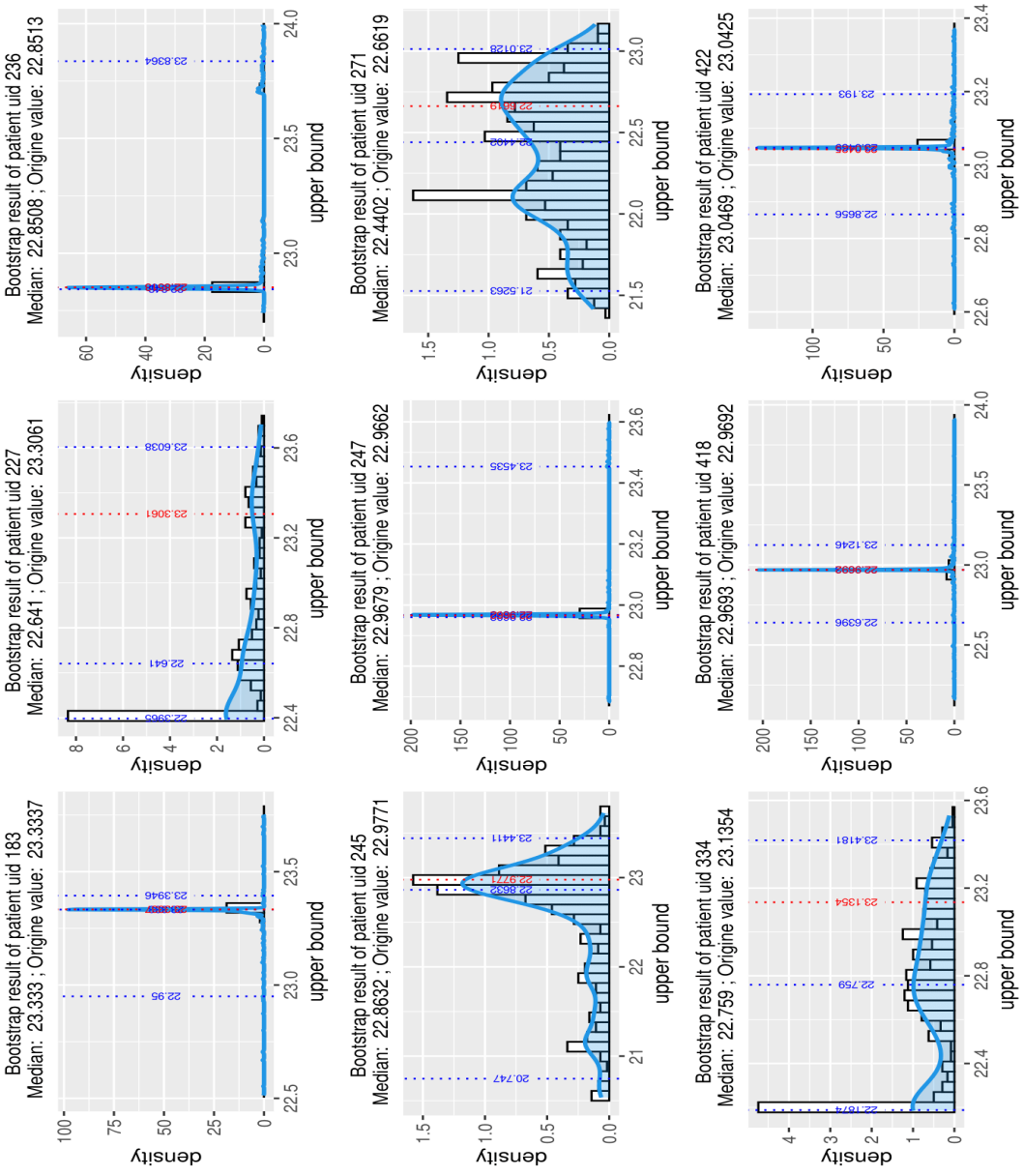


Figure 30: Upper bound page 2 for the right hand data. Histogram of the upper bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

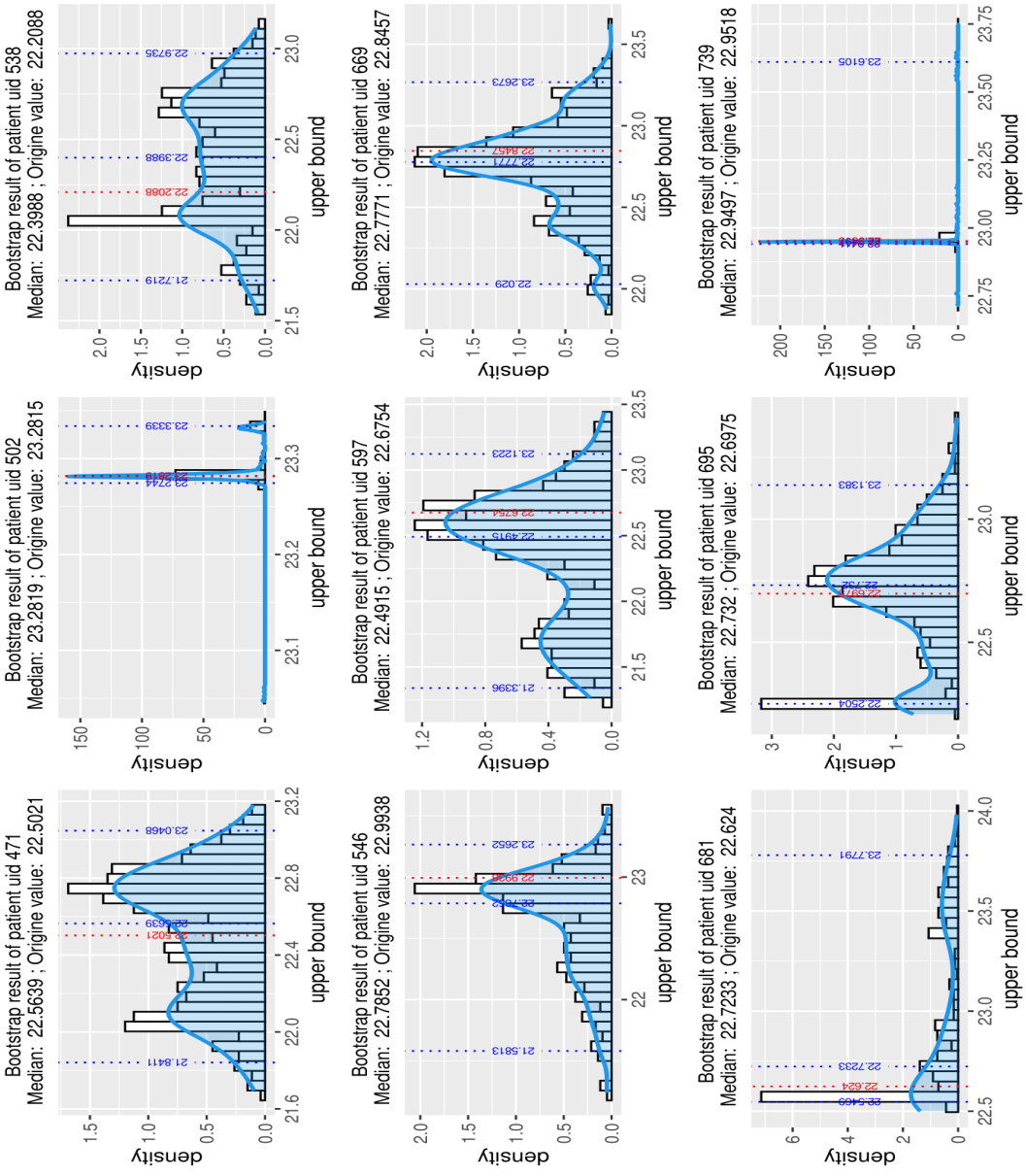


Figure 31: Upper bound page 3 for the right hand data. Histogram of the upper bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

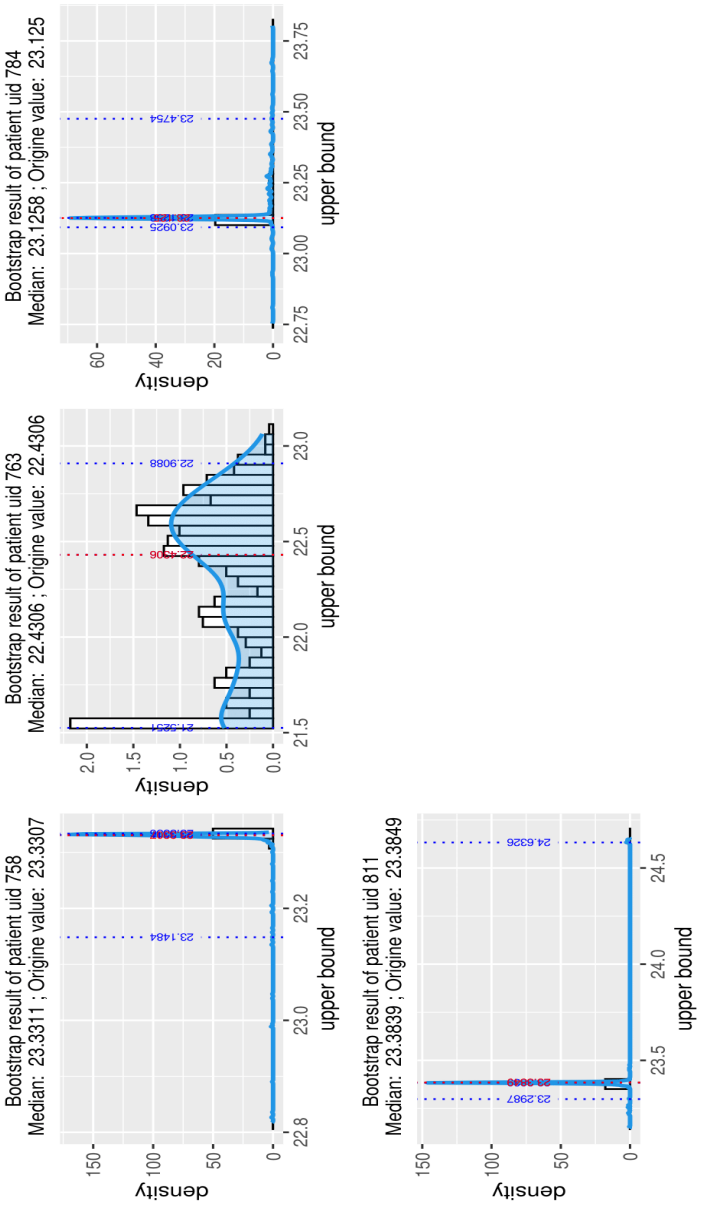


Figure 32: Upper bound page 4 for the right hand data. Histogram of the upper bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

E Confidence Interval of IRI for right foot data

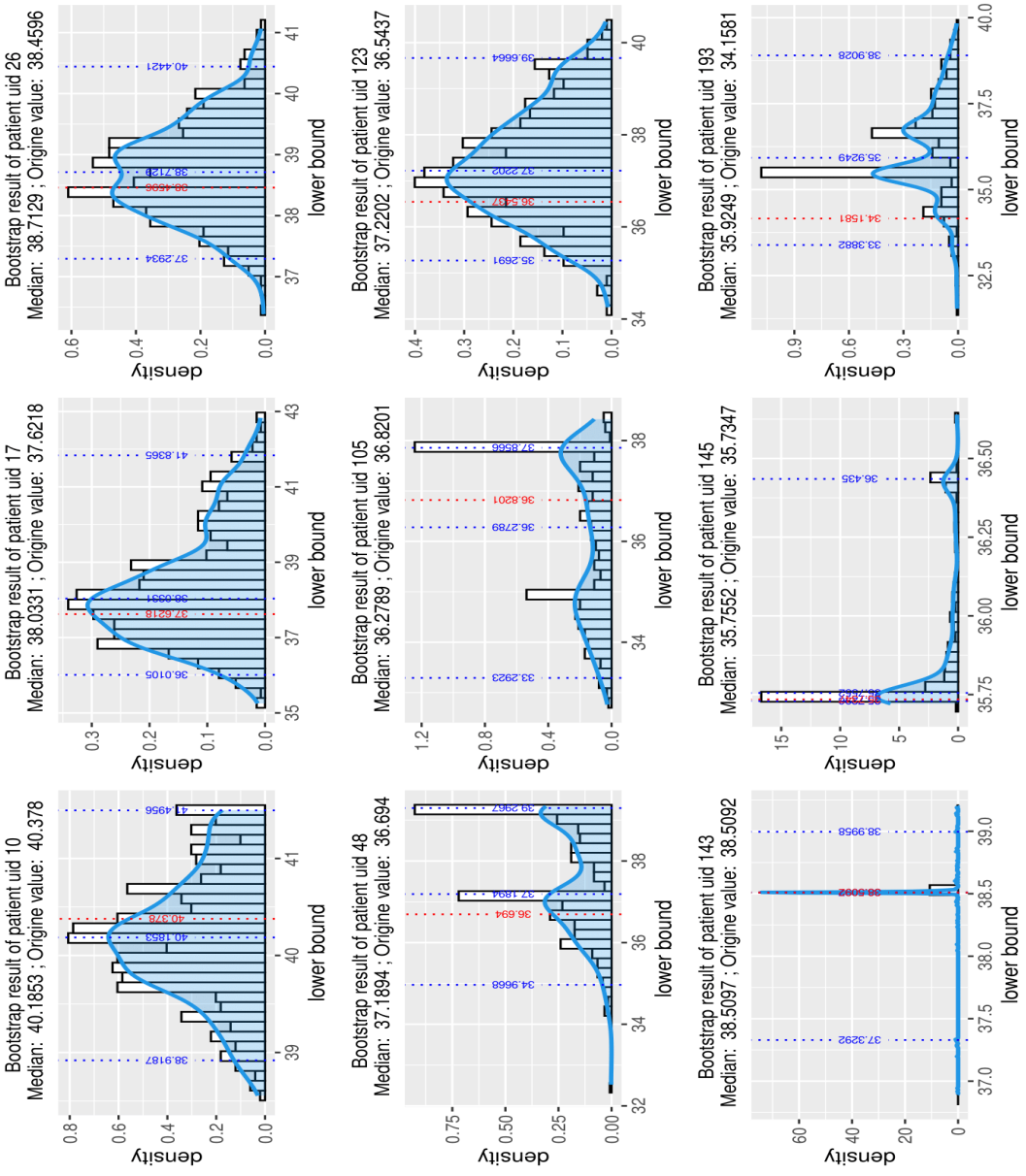


Figure 33: Lower bound page 1 for the right foot data. Histogram of the lower bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

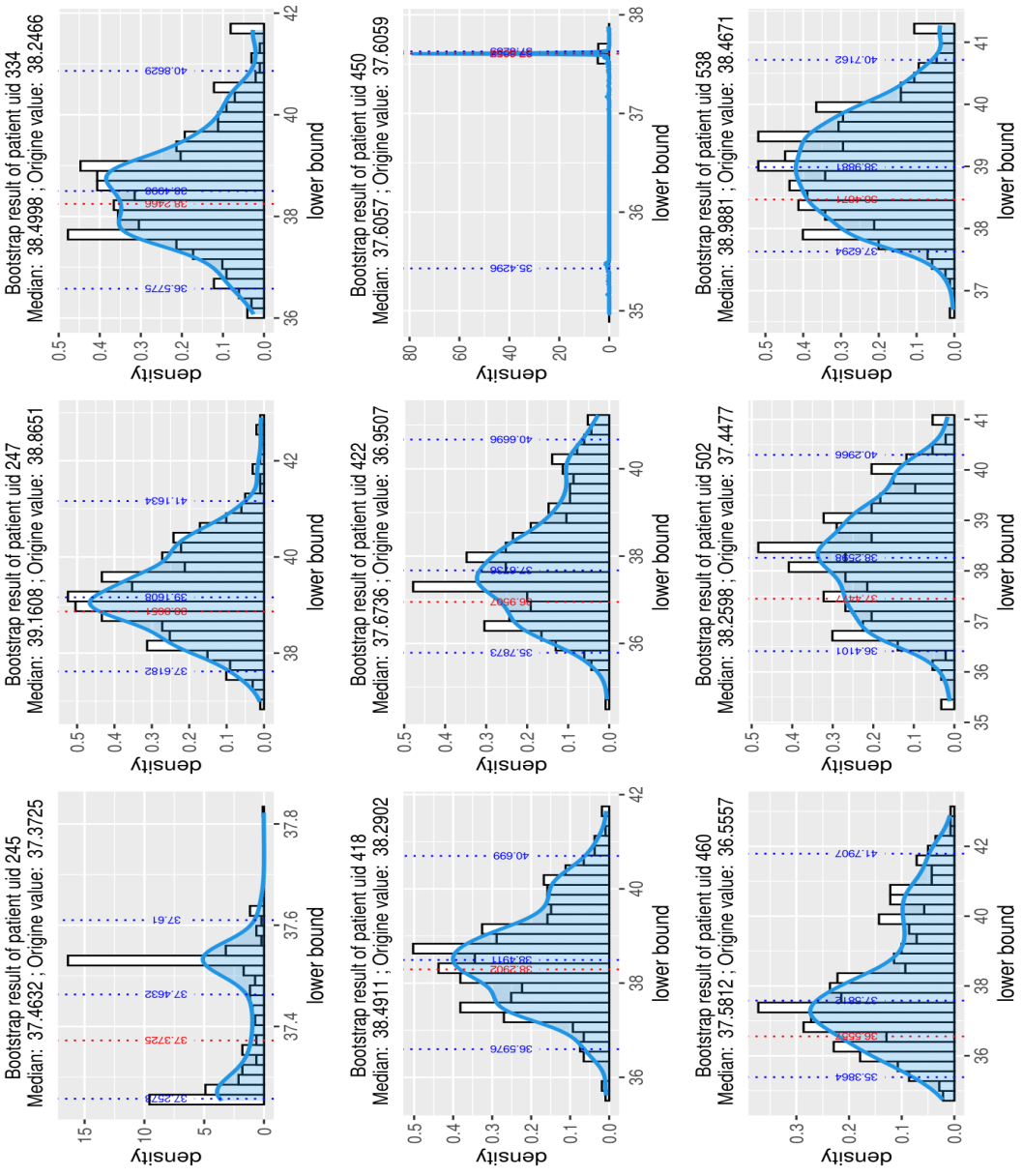


Figure 34: Lower bound page 2 for the right foot data. Histogram of the lower bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

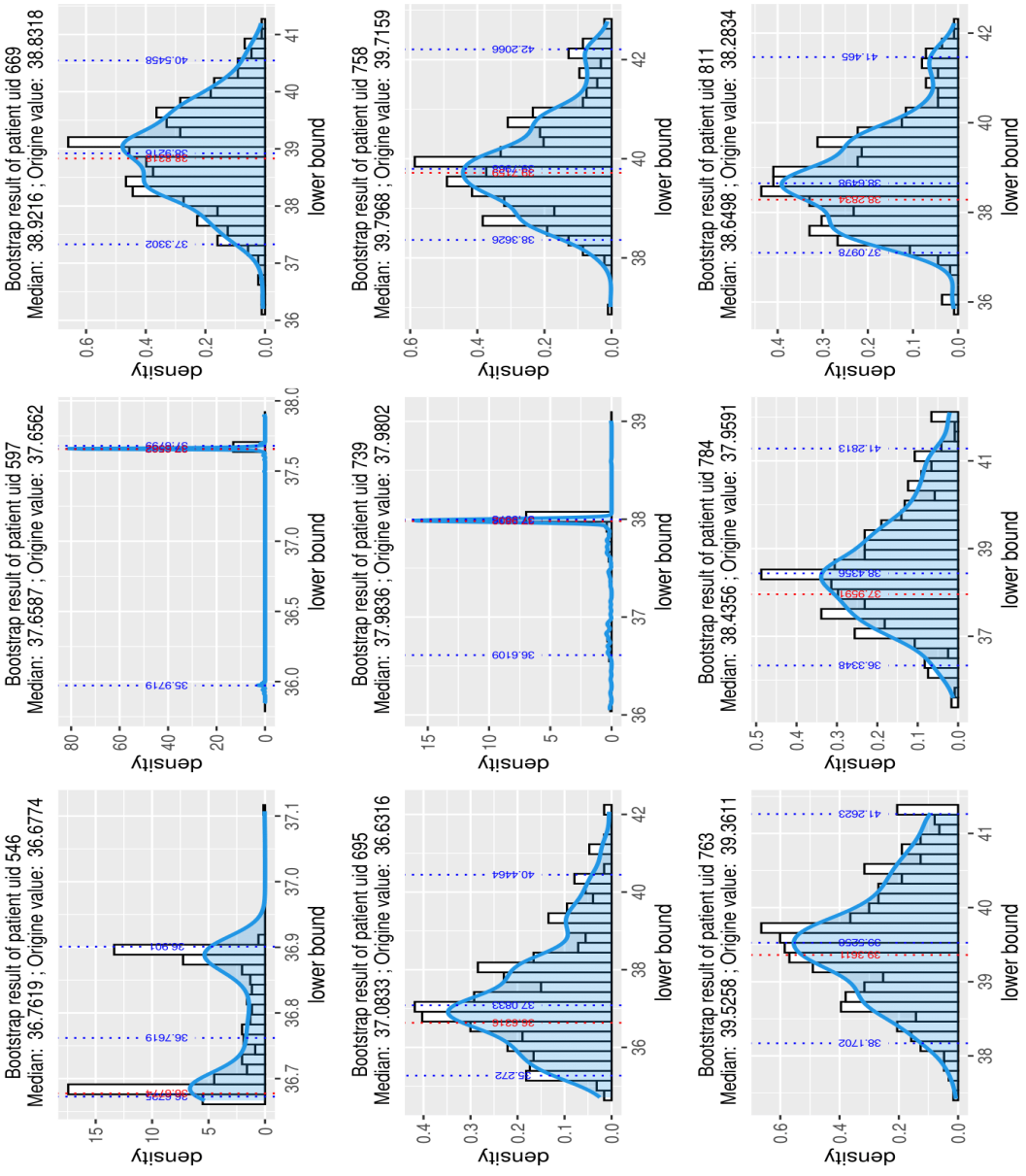


Figure 35: Lower bound page 3 for the right foot data. Histogram of the lower bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

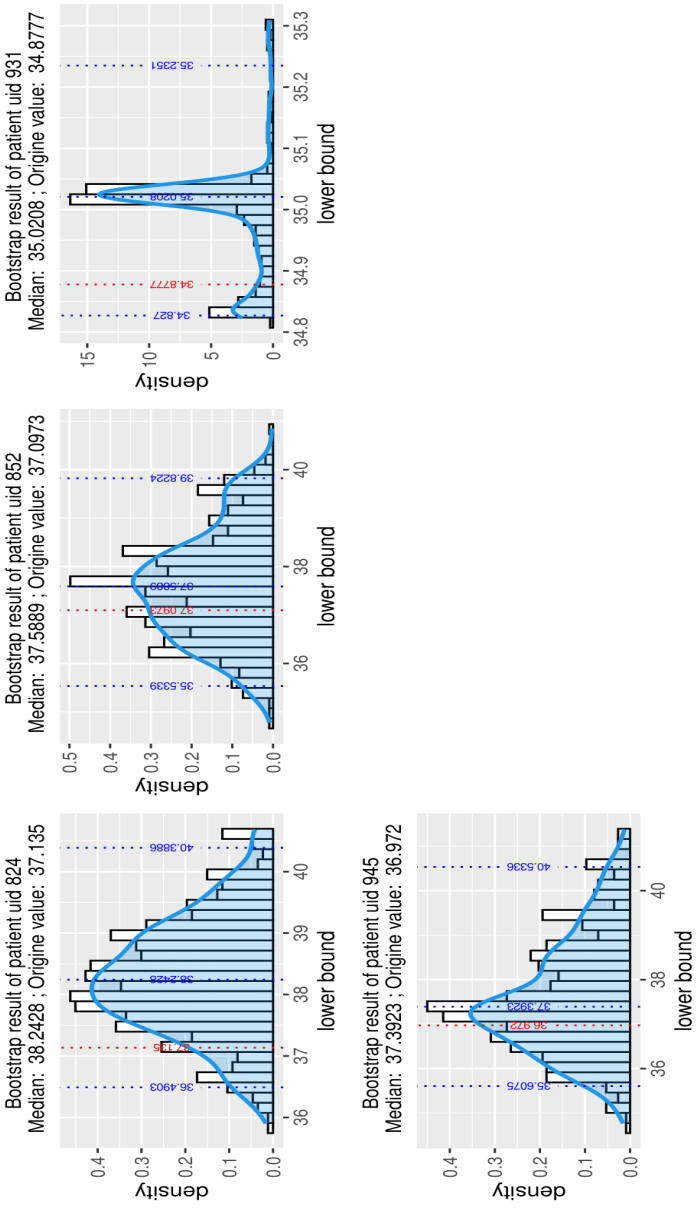


Figure 36: Lower bound page 4 for the right foot data. Histogram of the lower bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

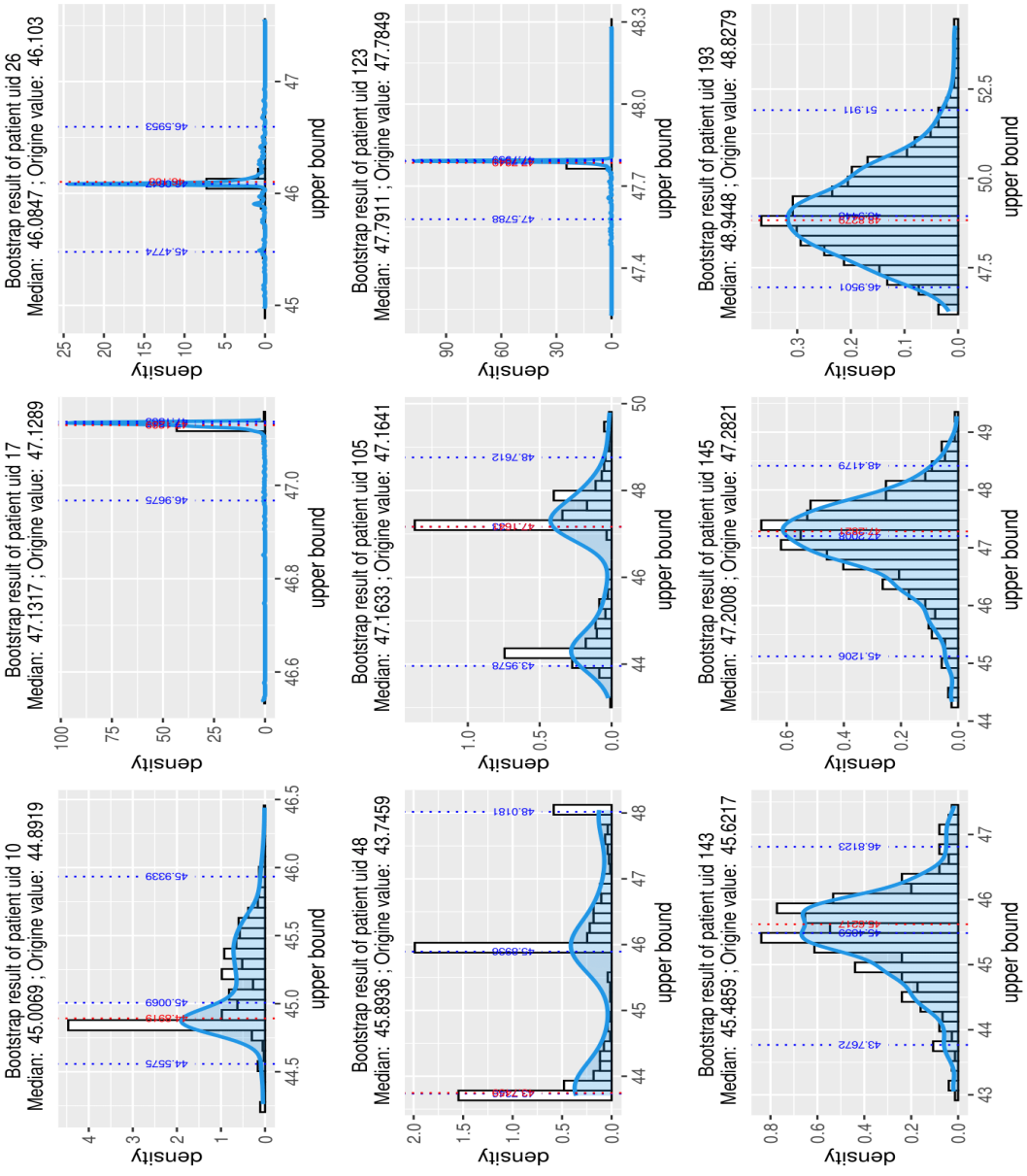


Figure 37: Upper bound page 1 for the right foot data. Histogram of the upper bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

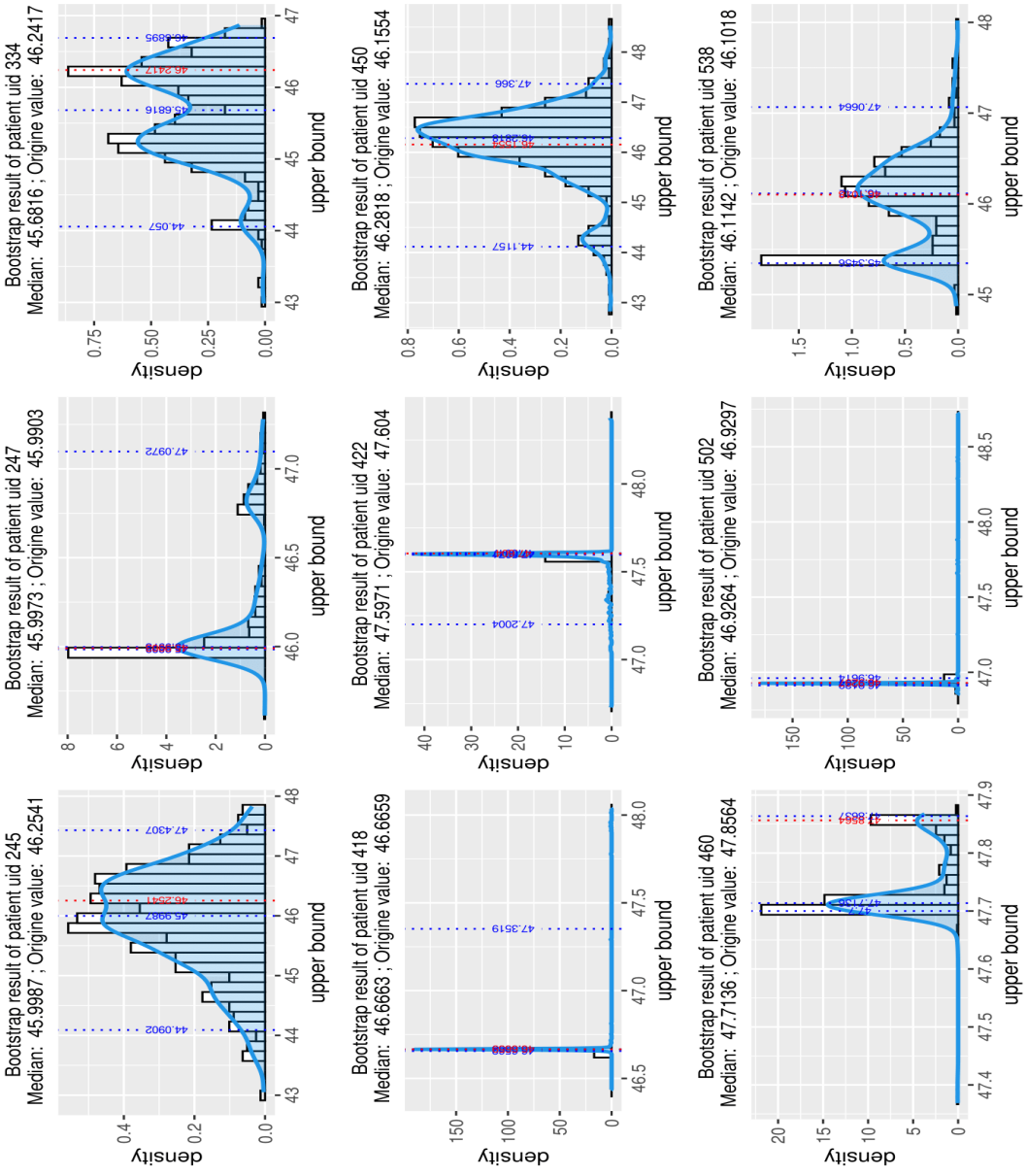


Figure 38: Upper bound page 2 for the right foot data. Histogram of the upper bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

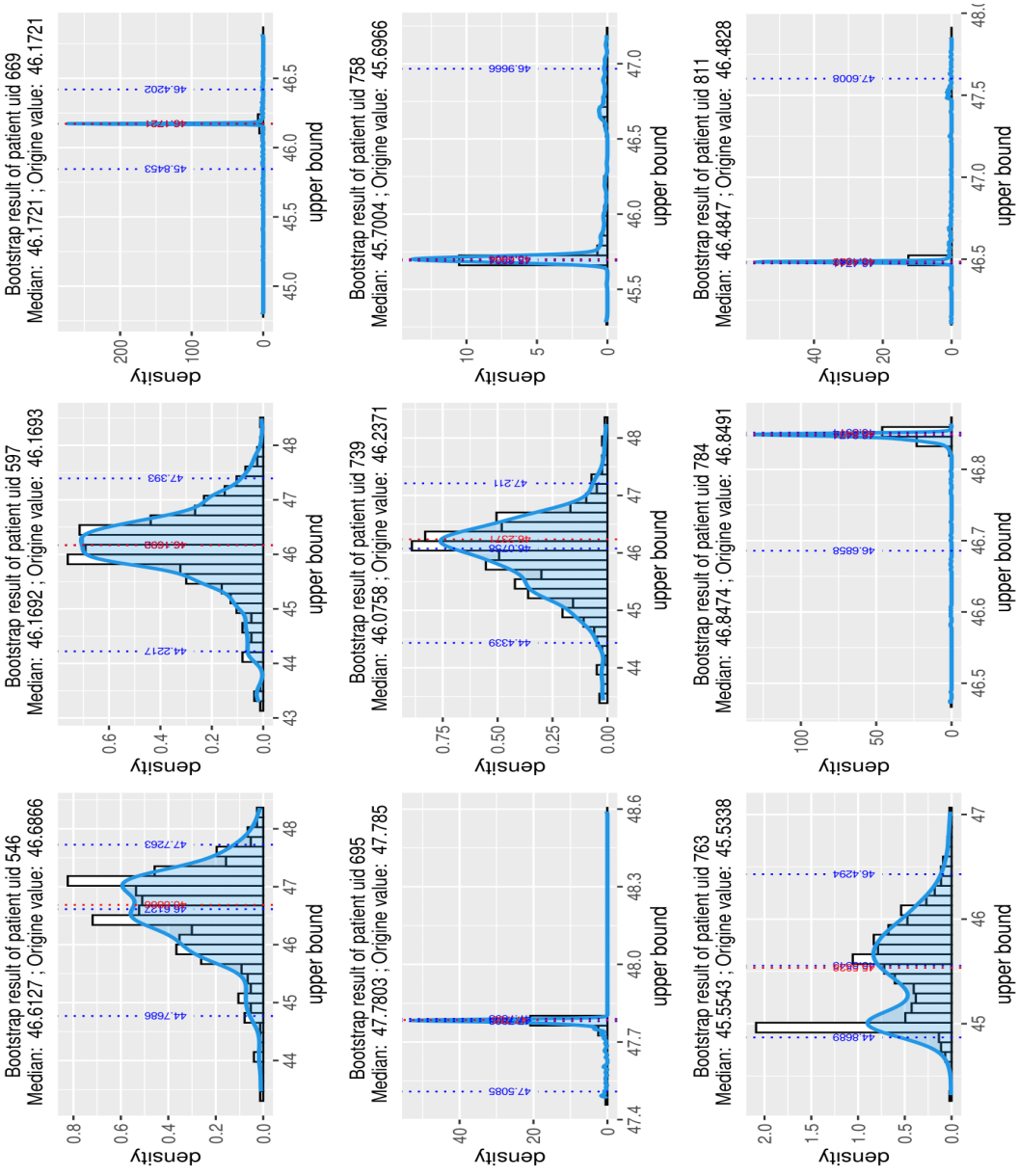


Figure 39: Upper bound page 3 for the right foot data. Histogram of the upper bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

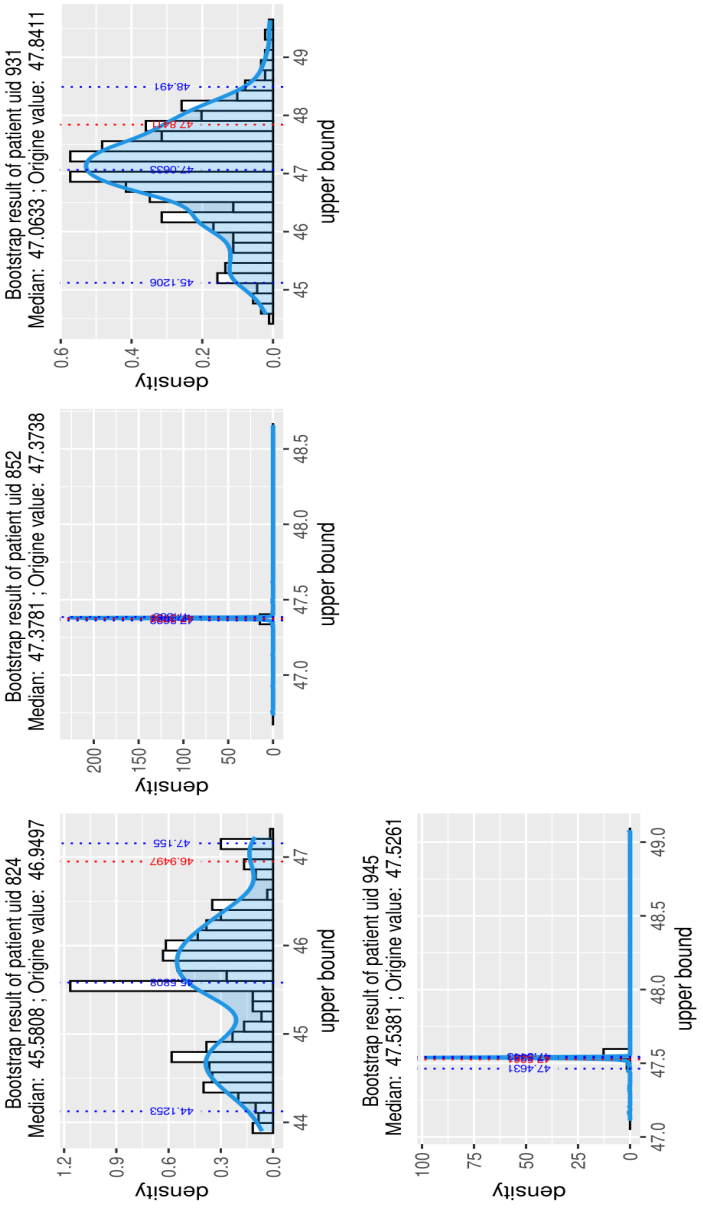


Figure 40: Upper bound page 4 for the right foot data. Histogram of the upper bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

F Confidence Interval of IRI for left hand data

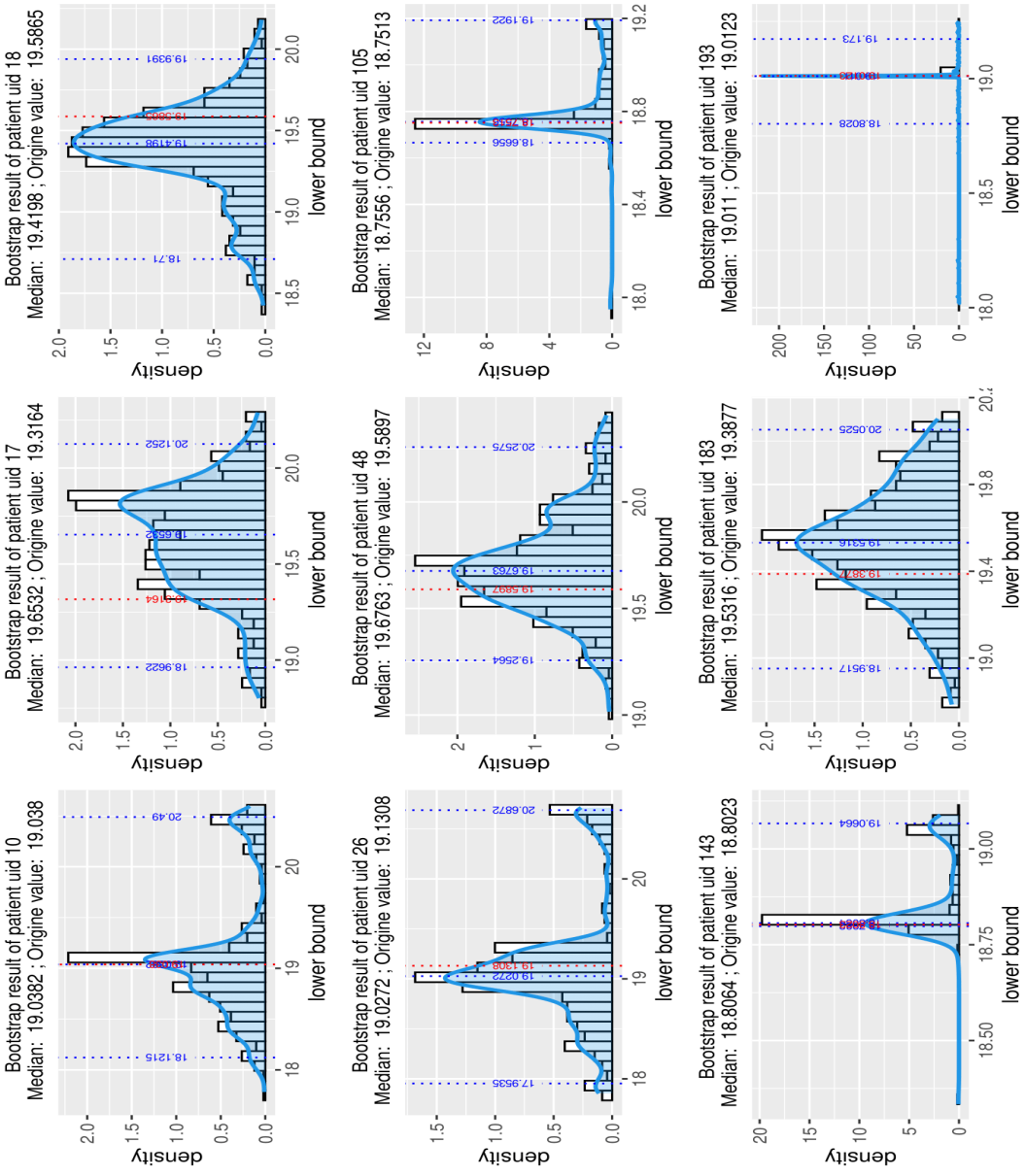


Figure 41: Lower bound page 1 for the left hand data. Histogram of the lower bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

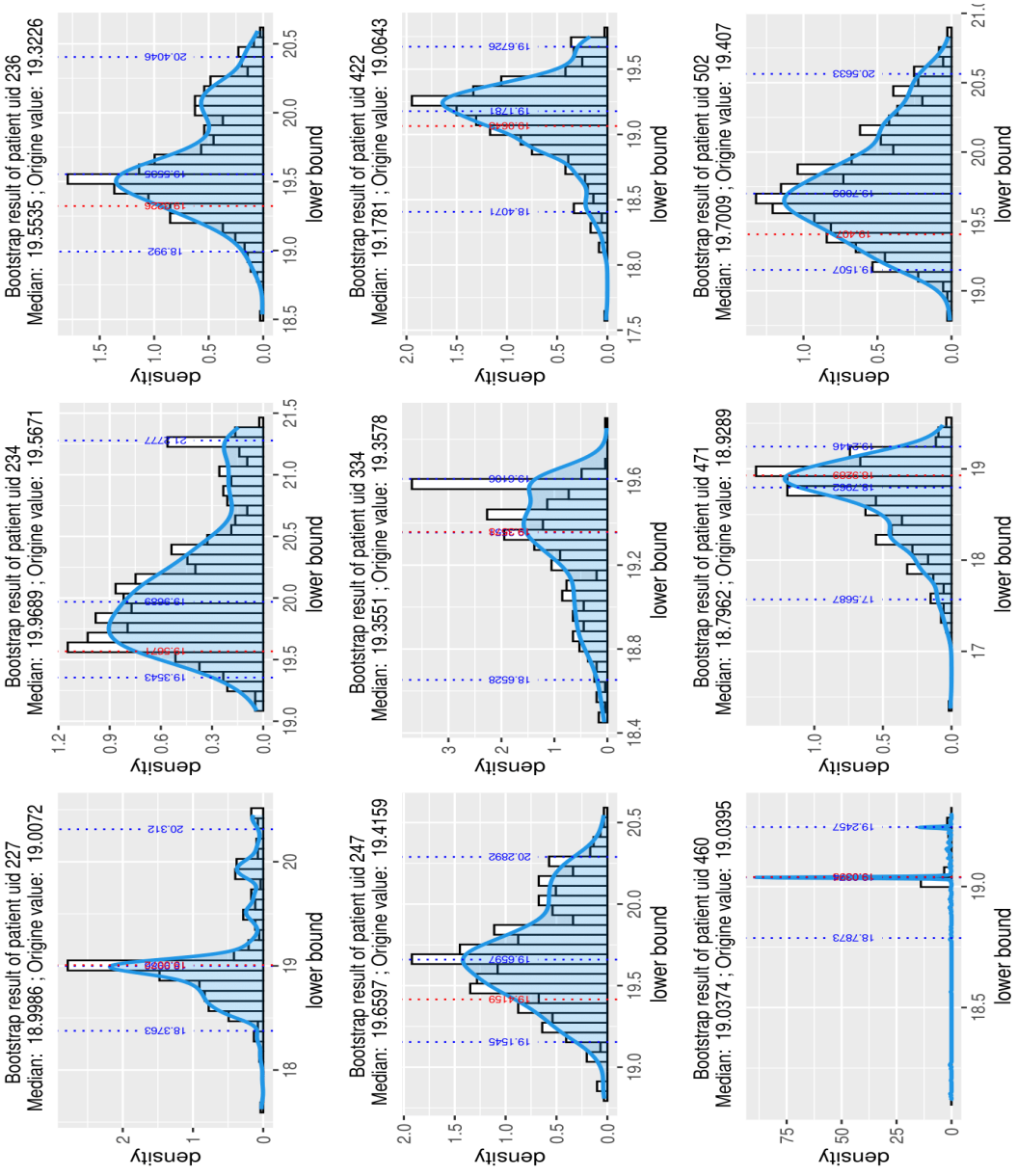


Figure 42: Lower bound page 2 for the left hand data. Histogram of the lower bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

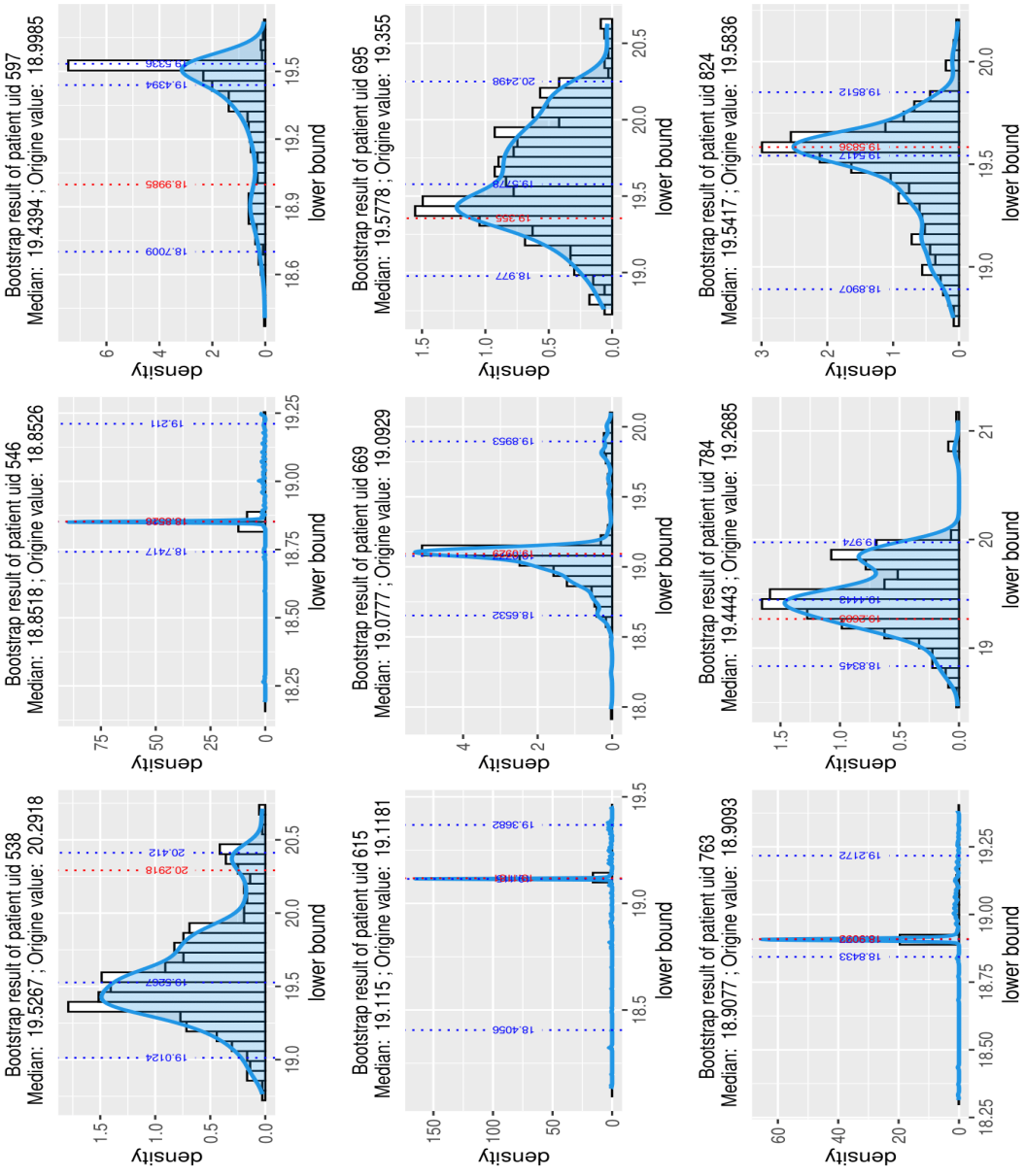


Figure 43: Lower bound page 3 for the left hand data. Histogram of the lower bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

Figure 44: Lower bound page 4 for the left hand data. Histogram of the lower bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

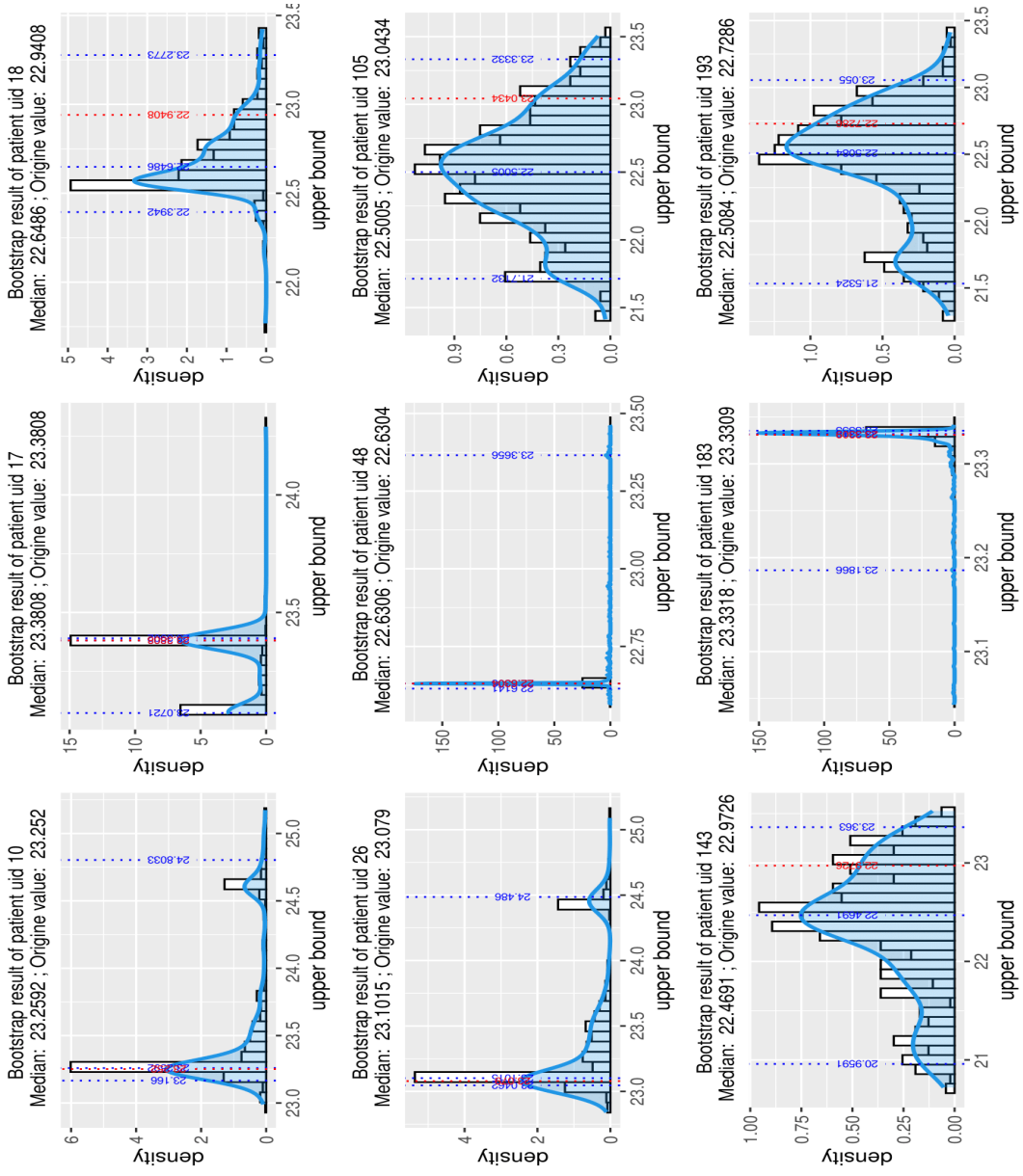


Figure 45: Upper bound page 1 for the left hand data. Histogram of the upper bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

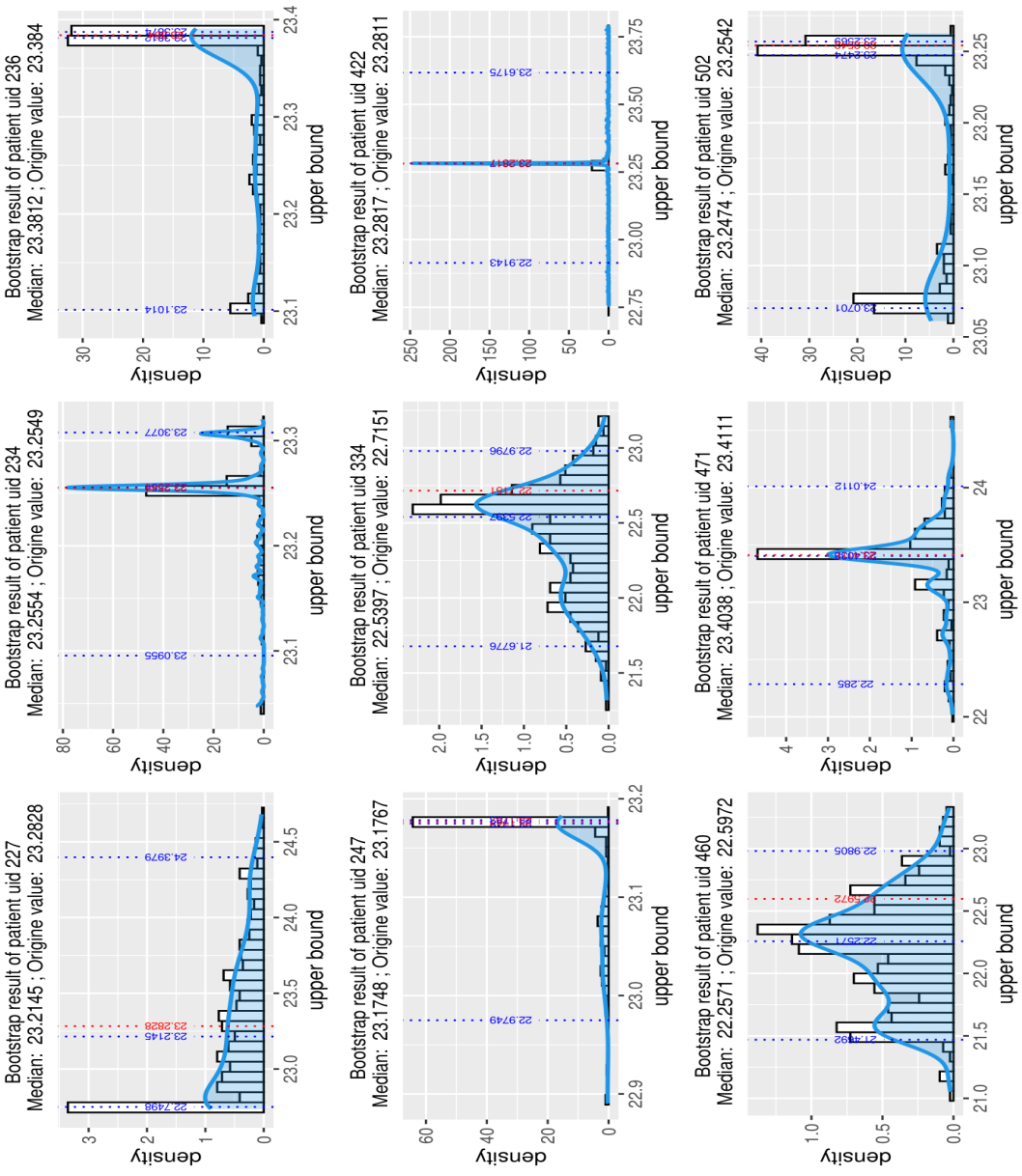


Figure 46: Upper bound page 2 for the left hand data. Histogram of the upper bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

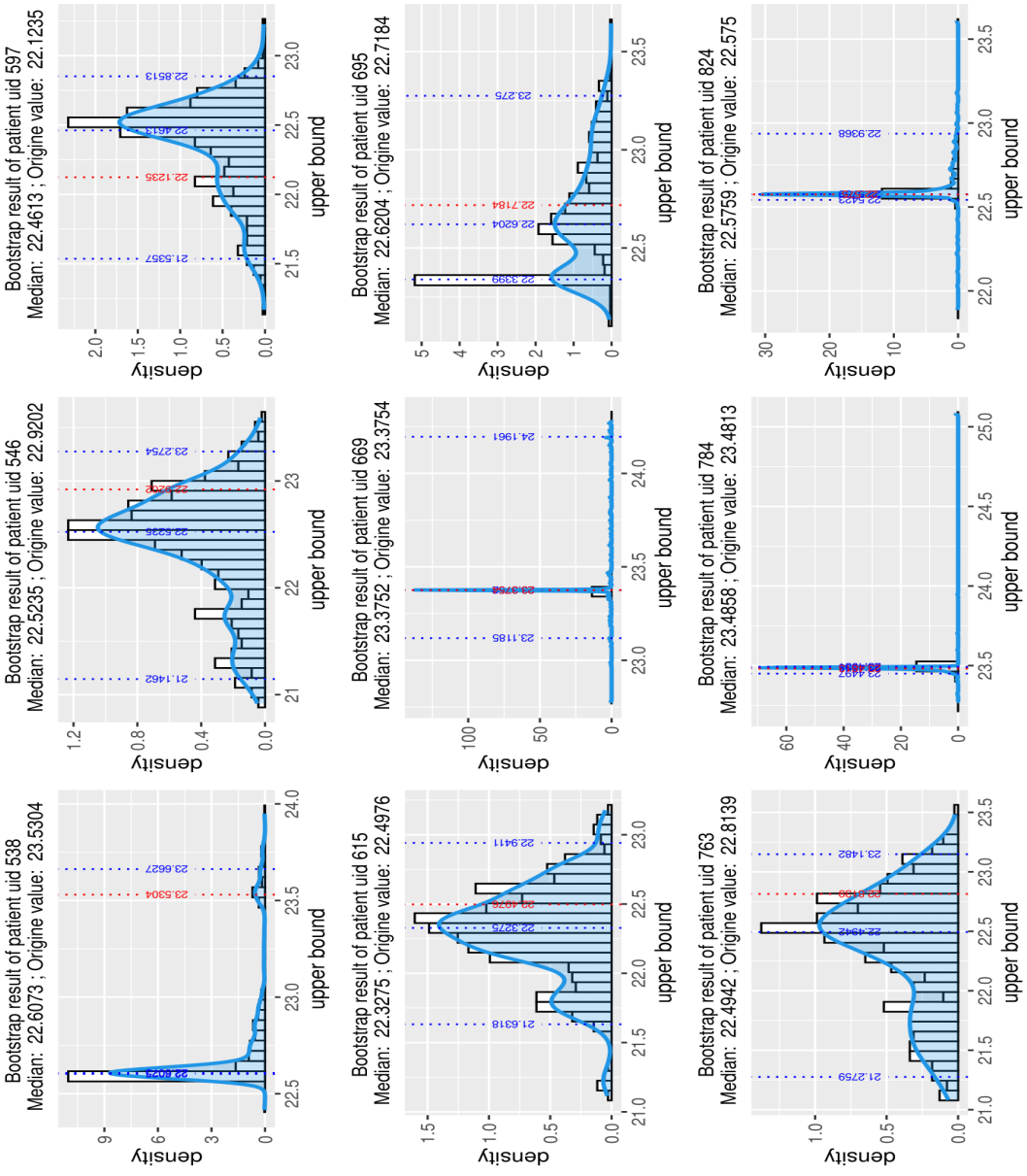


Figure 47: Upper bound page 3 for the left hand data. Histogram of the upper bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

Figure 48: Upper bound page 4 for the left hand data. Histogram of the upper bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

G Confidence Interval of IRI for left foot data

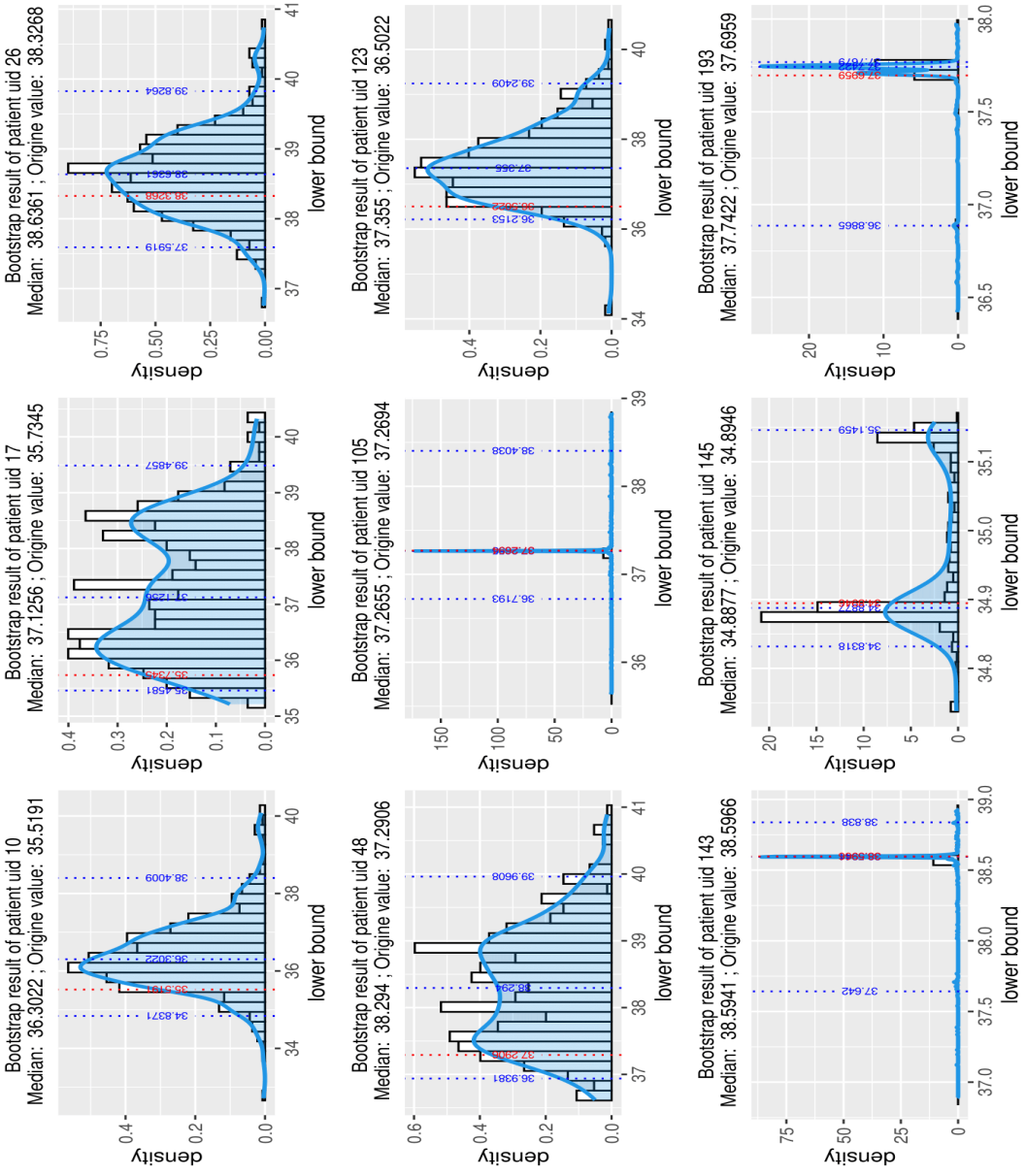


Figure 49: Lower bound page 1 for the left foot data. Histogram of the lower bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

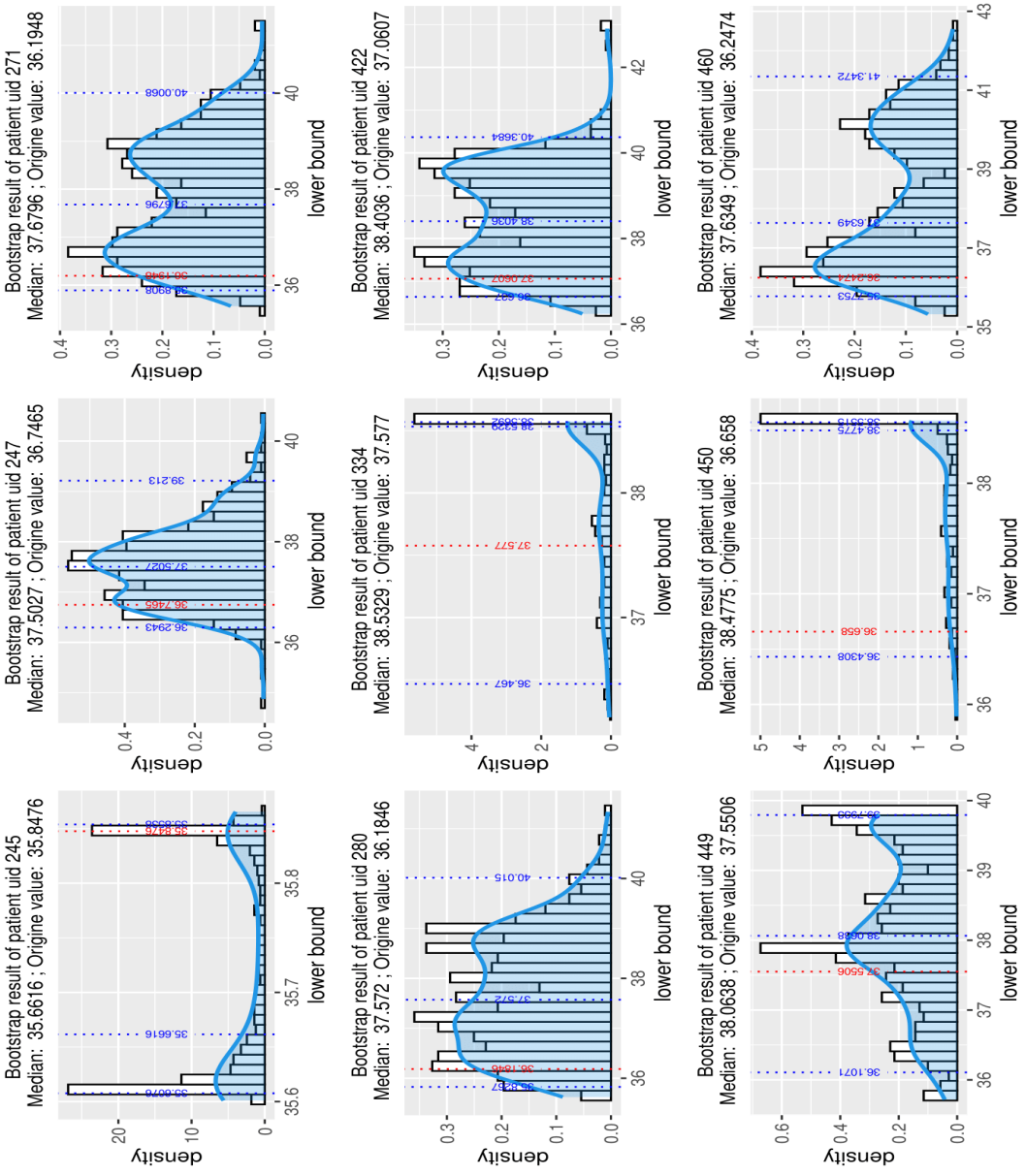


Figure 50: Lower bound page 2 for the left foot data. Histogram of the lower bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

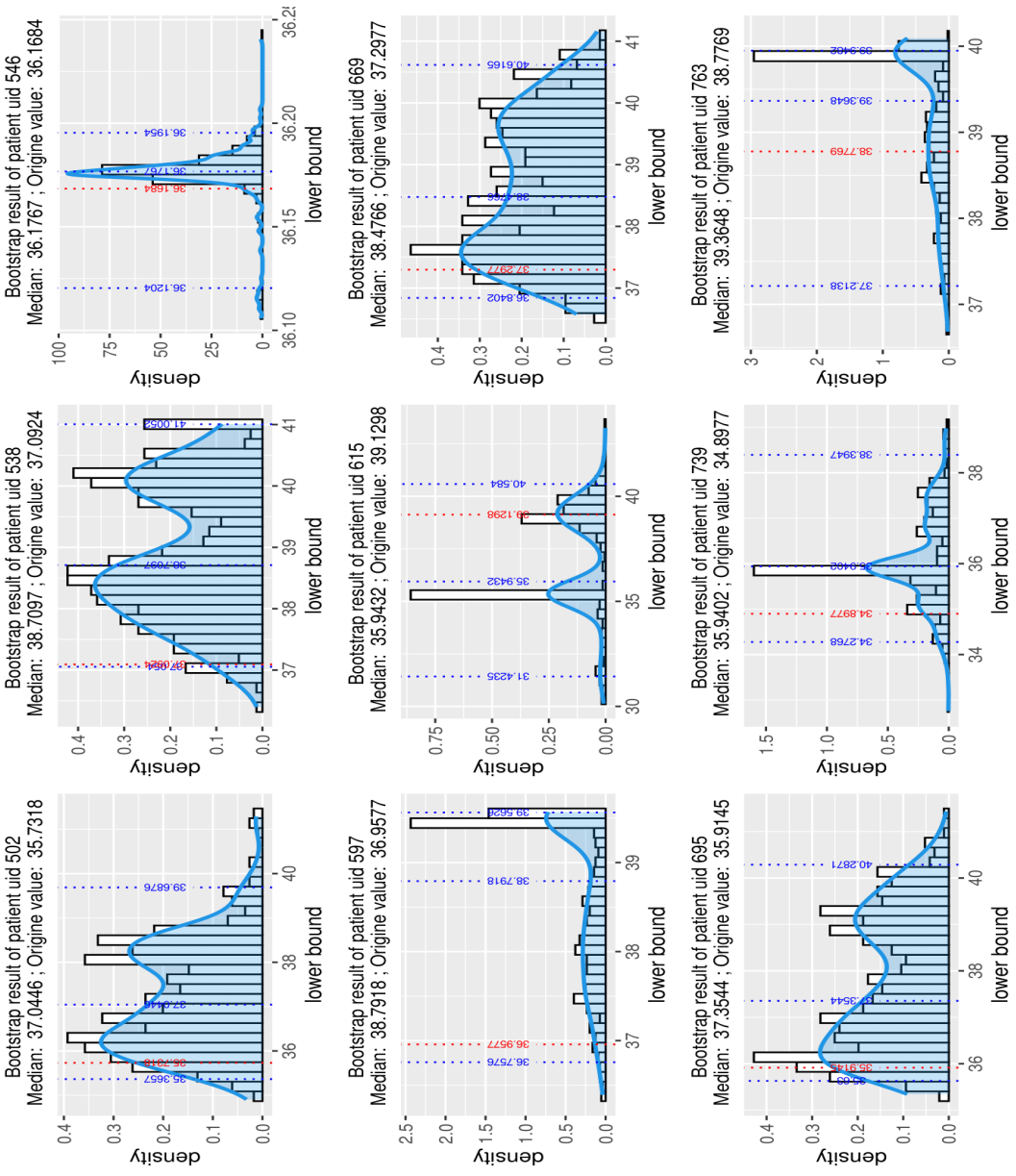


Figure 51: Lower bound page 3 for the left foot data. Histogram of the lower bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

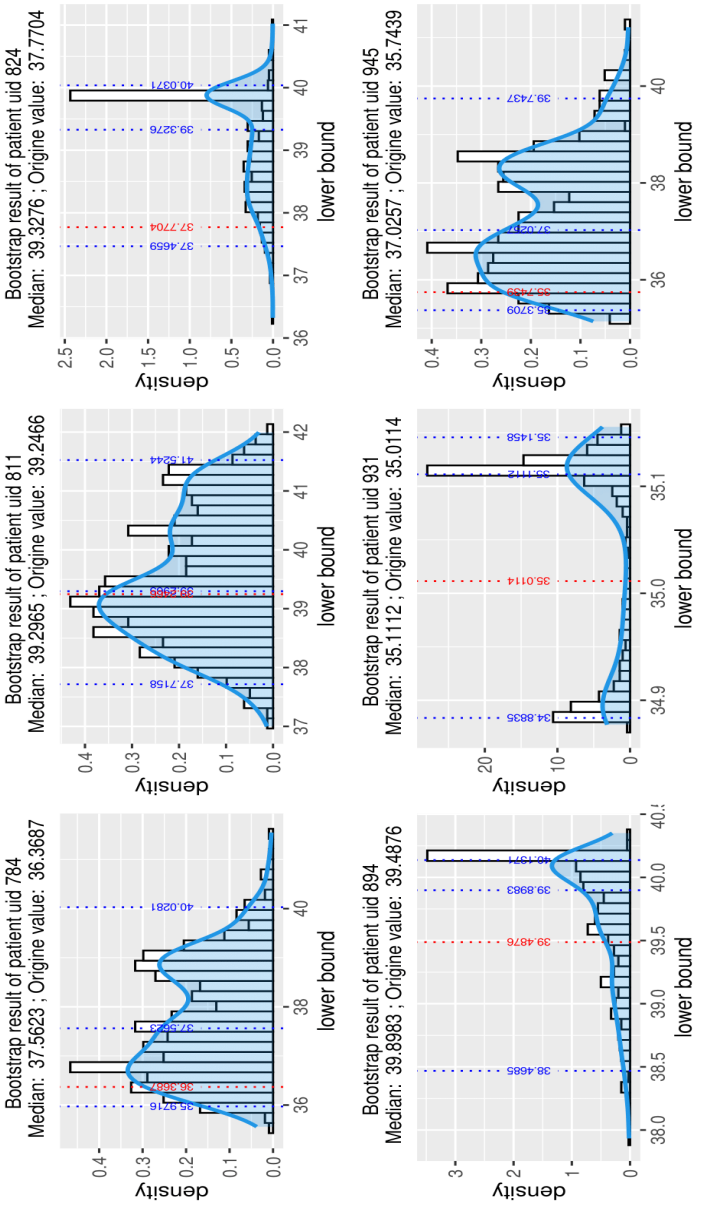


Figure 52: Lower bound page 4 for the left foot data. Histogram of the lower bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

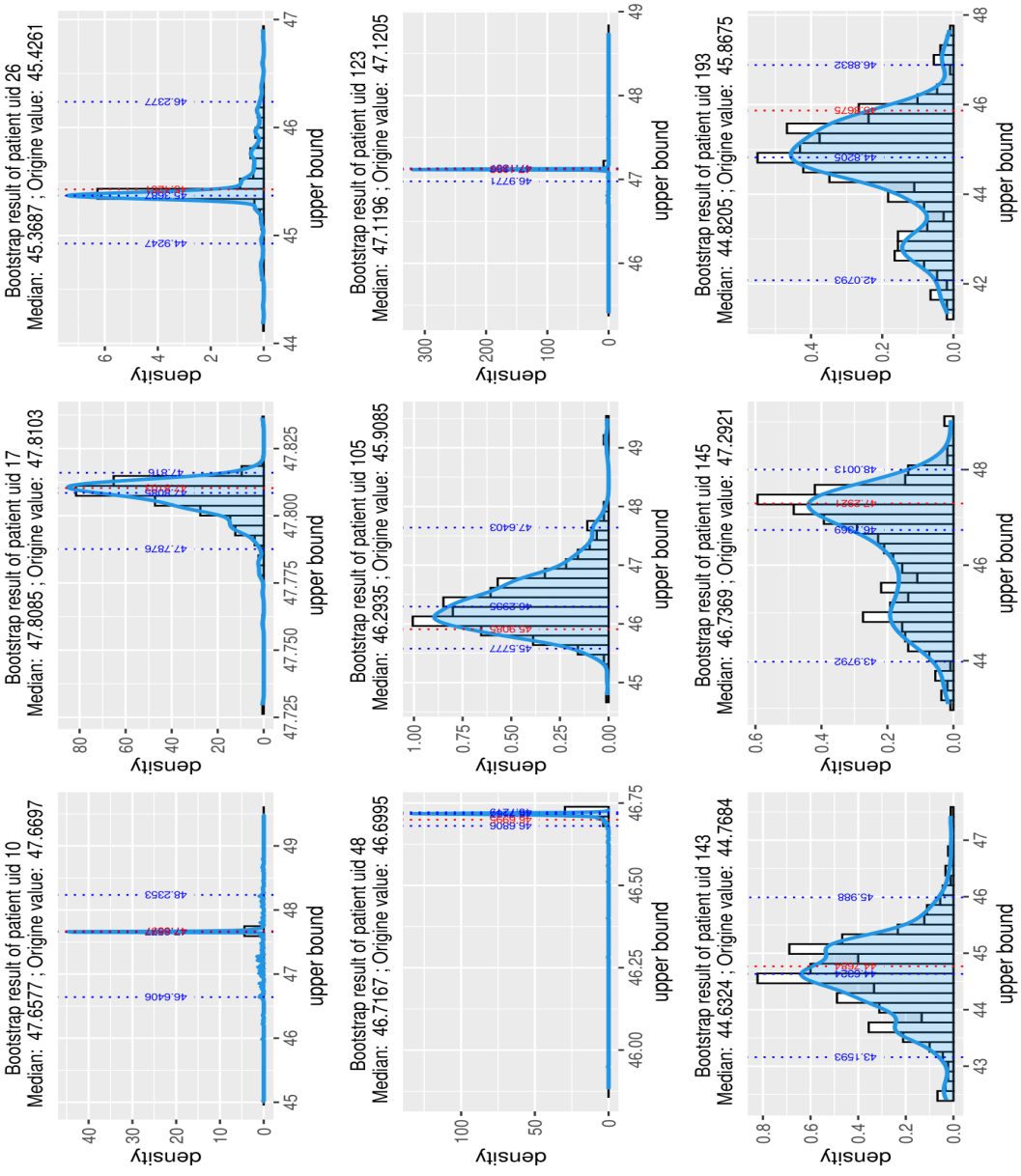


Figure 53: Upper bound page 1 for the left foot data. Histogram of the upper bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

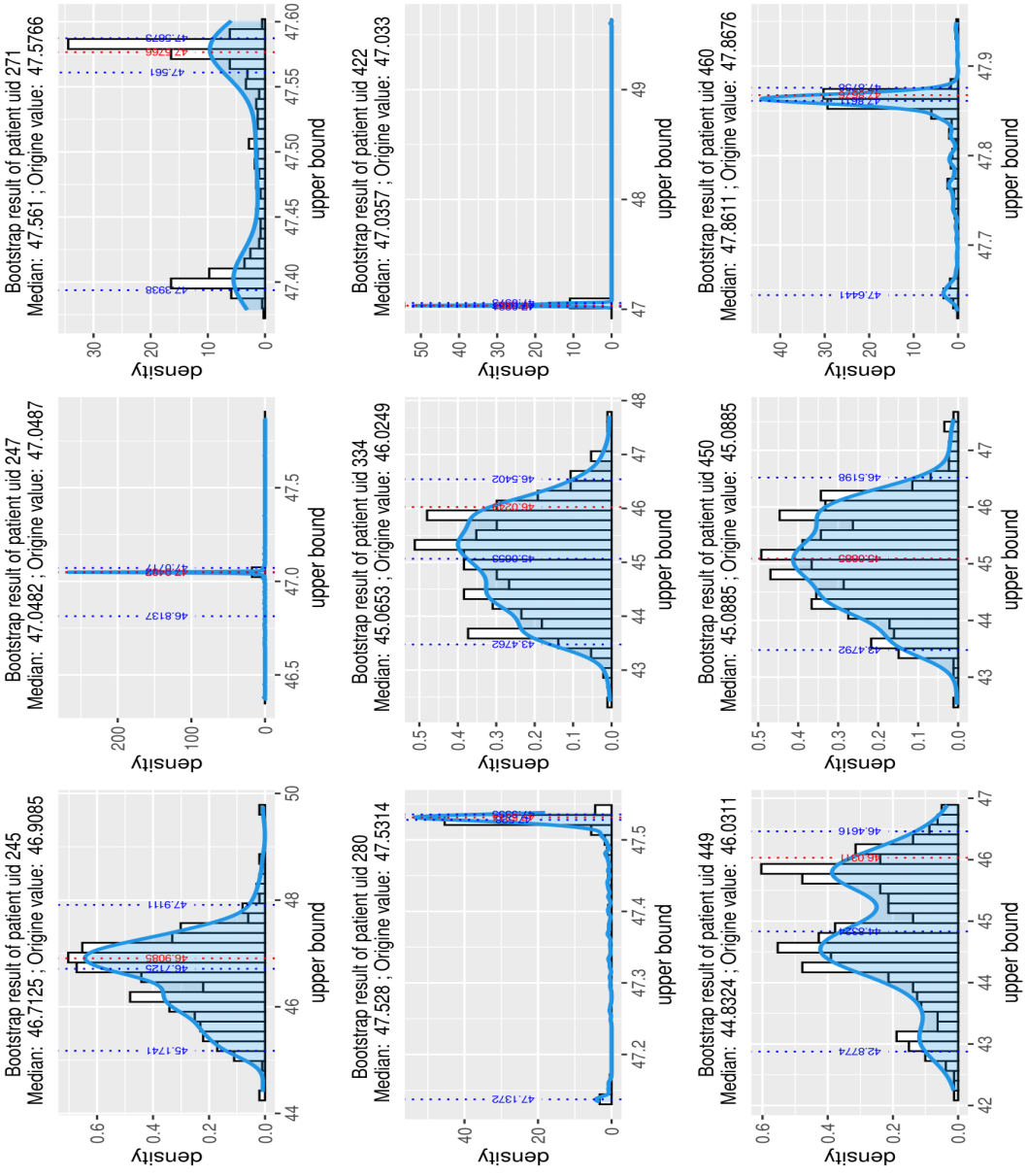


Figure 54: Upper bound page 2 for the left foot data. Histogram of the upper bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

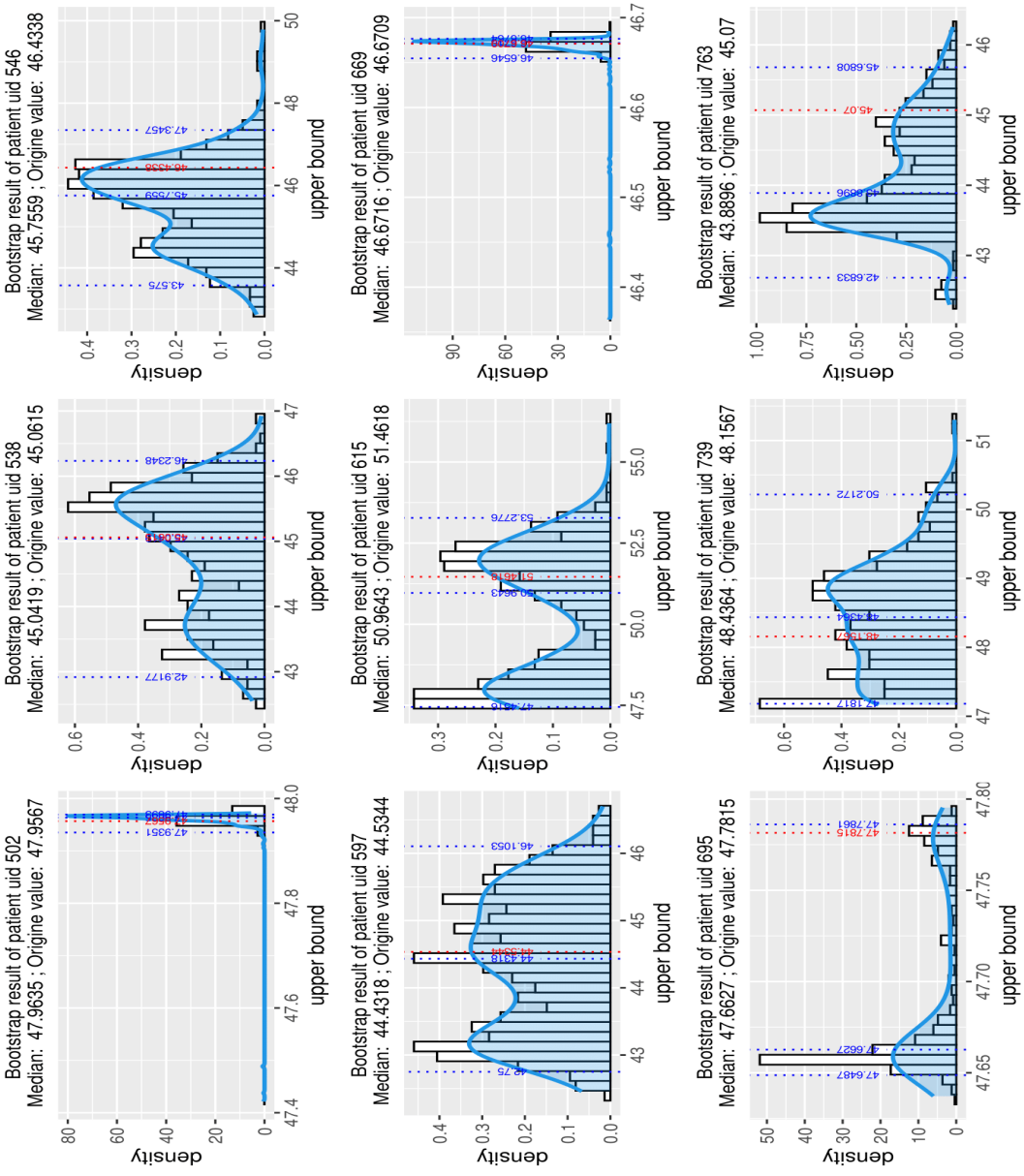


Figure 55: Upper bound page 3 for the left foot data. Histogram of the upper bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

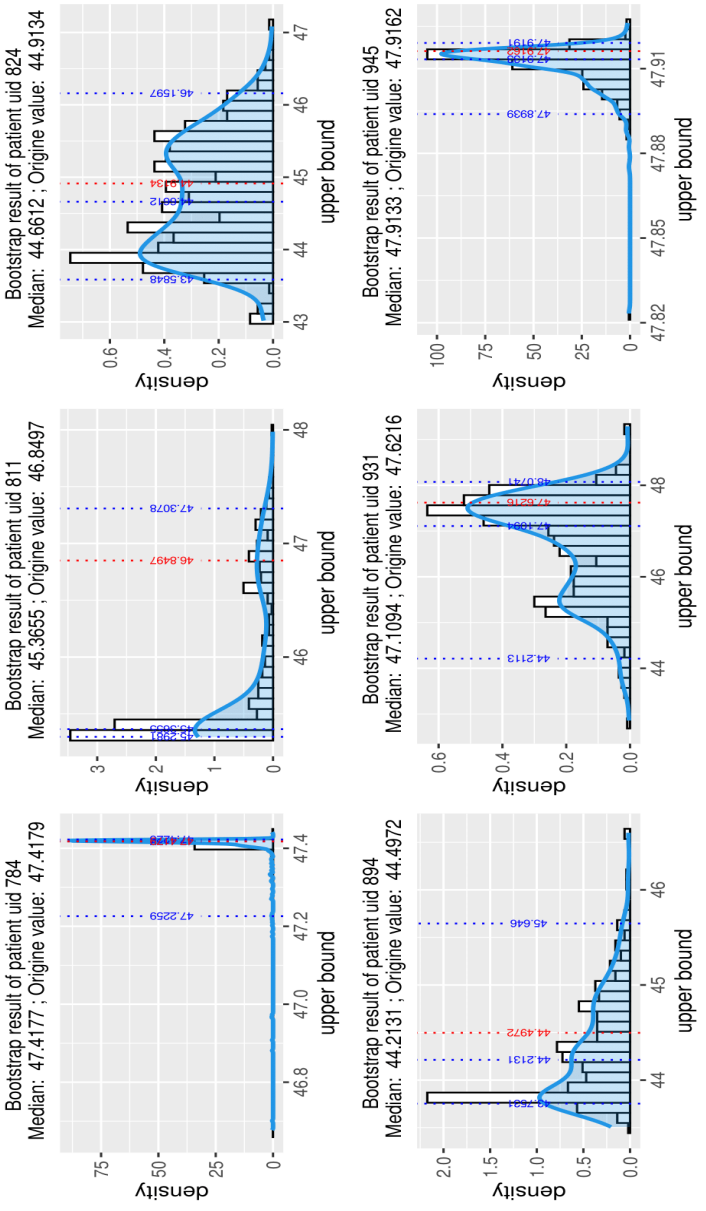
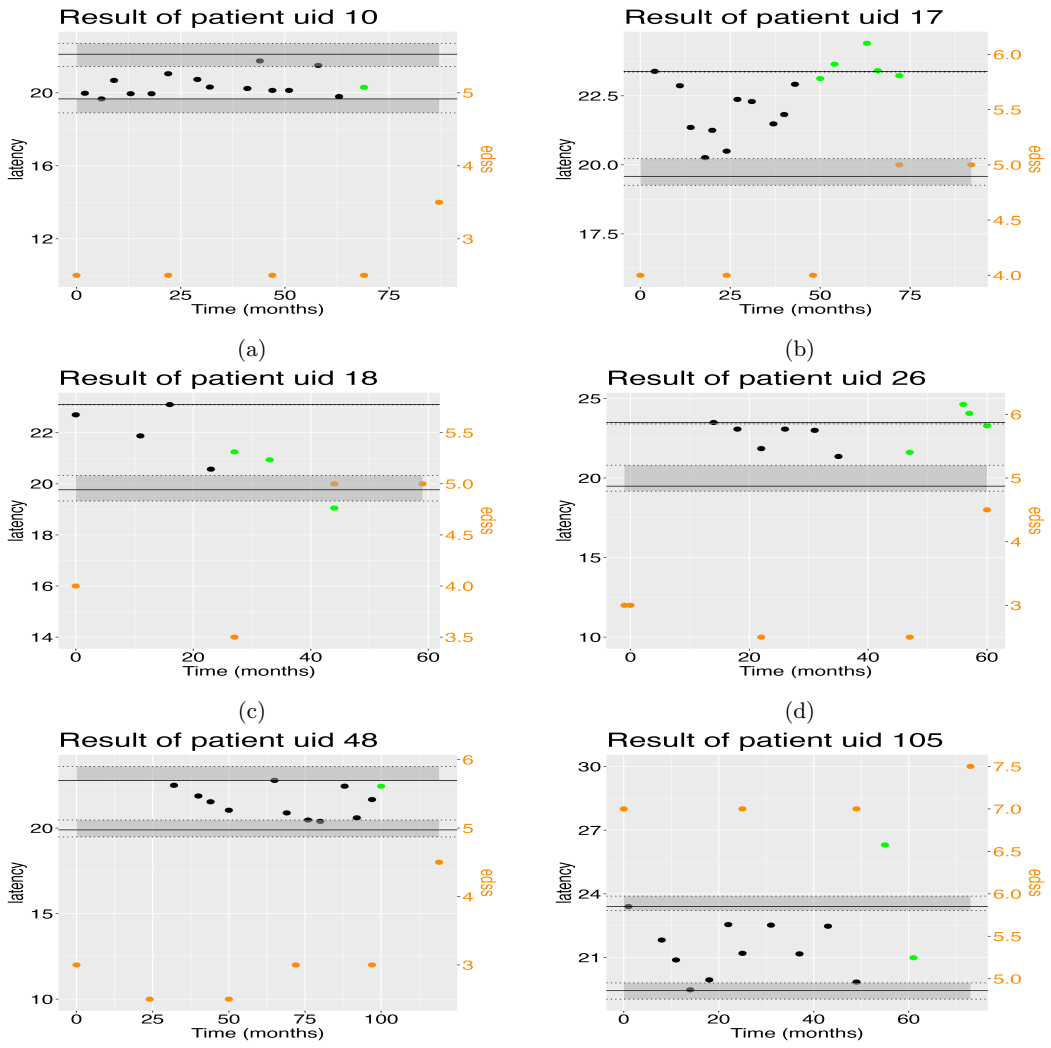


Figure 56: Upper bound page 4 for the left foot data. Histogram of the upper bound of each ID is shown. The red dashed line is the lower bound from the origin dataset. The dashed blue lines are the 2.5th, 50th, and 97.5th percentile of the 500 Bootstrap values. In most cases, the original value and the median are quite close.

H Confidence Interval of IRI and EDSS for right hand data



(a) (b) (c) (d) (e) (f)

Figure 57: A subset of PwMS from the right hand data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points present the EDSS score. The progression of MS can be detected earlier using IRI on latency measurement than EDSS score.

(a) No data available to detect the progression based on IRI

(b) Progression can be detected at 72 months based on EDSS score, and at 54 months based on IRI.

(c) Progression cannot be detected sooner by IRI than by EDSS score

(d) Progression can be detected at 60 months based on EDSS score, and at 56 months based on IRI.

(e) No data available to detect the progression based on IRI.

(f) Progression can be detected at 73 months based on EDSS score, and at 55 months based on IRI.

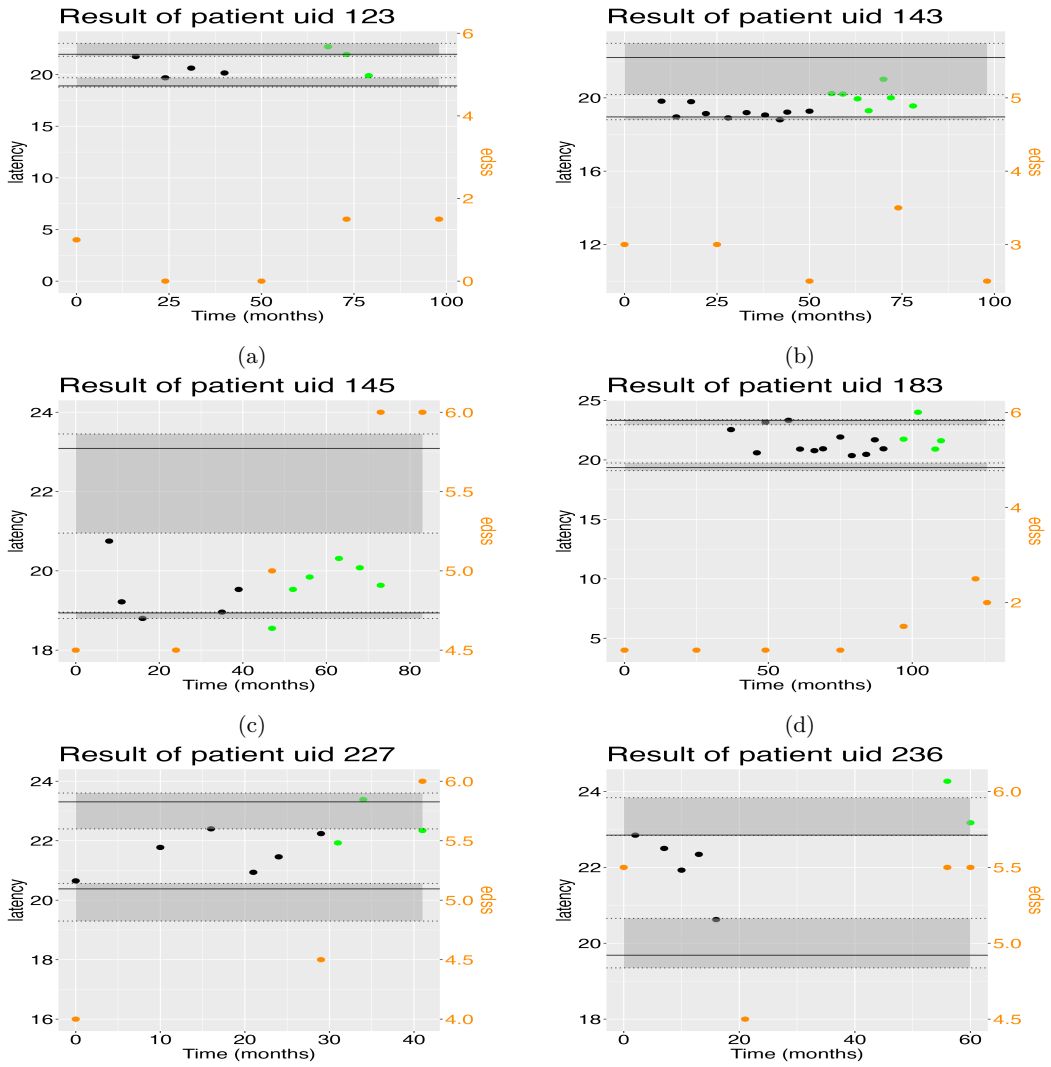


Figure 58: A subset of PwMS from the right hand data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points present the EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression can be detected at 73 months based on EDSS score, and at 68 months based on IRI.
(b) Progression can be detected at 74 months based on EDSS score, and at 56 months based on IRI.
(c) Progression cannot be detected earlier by IRI than EDSS score due to no data vailable during the time interval of 2 years
(d) Progression can be detected at 122 months based on EDSS score, and at 102 months based on IRI.
(e) Progression can be detected at 41 months based on EDSS score, and at 34 months based on IRI.
(f) Progression cannot be detected earlier by IRI than EDSS score due to no data available during the time interval of 2 years

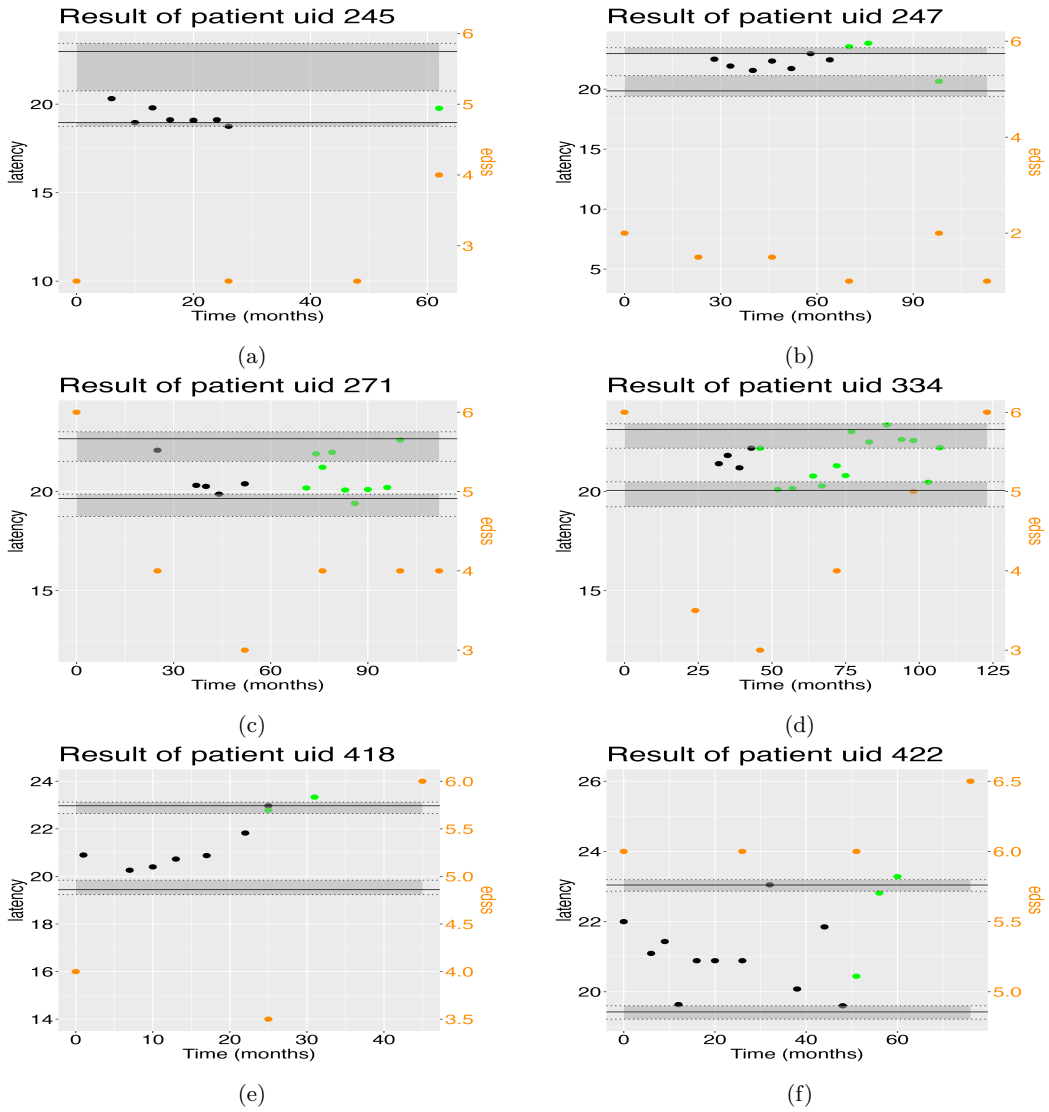


Figure 59: A subset of PwMS from the right hand data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points present the EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression cannot be detected earlier by IRI than EDSS score due to no data available during the time intervla of 2 years
(b) Progression can be detected at 98 months based on EDSS score, and at 70 months based on IRI.
(c) Progression cannot be detected earlier by IRI than EDSS score due to no data available during the time intervla of 2 years.
(d) Progression can be detected at 72 months based on EDSS score, and at 46 months based on IRI.
(e) Progression can be detected at 45 months based on EDSS score, and at 25 months based on IRI.
(f) Progression can be detected at 76 months based on EDSS score, and at 56 months based on IRI.

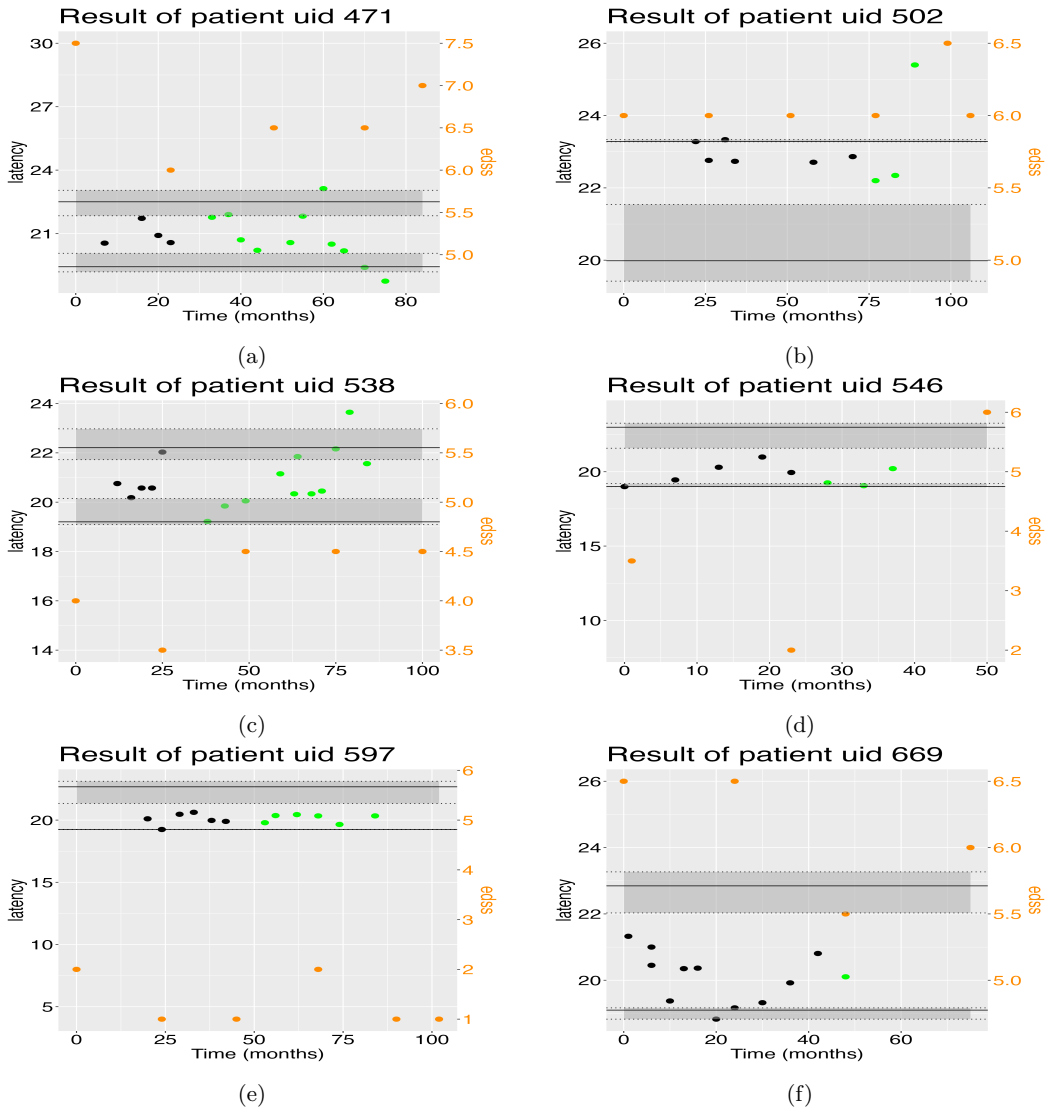


Figure 60: A subset of PwMS from the right hand data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression can be detected at 48 months based on EDSS score, and at 33 months based on IRI.
(b) Progression can be detected at 99 months based on EDSS score, and at 89 months based on IRI.
(c) Progression can be detected at 49 months based on EDSS score, and at 38 months based on IRI.
(d) Progression can be detected at 50 months based on EDSS score, and at 28 months based on IRI.
(e) Progression cannot detected earlier by IRI than by EDSS score
(f) Progression cannot be detected earlier by IRI than EDSS score due to no data available during the time intervla of 2 years.

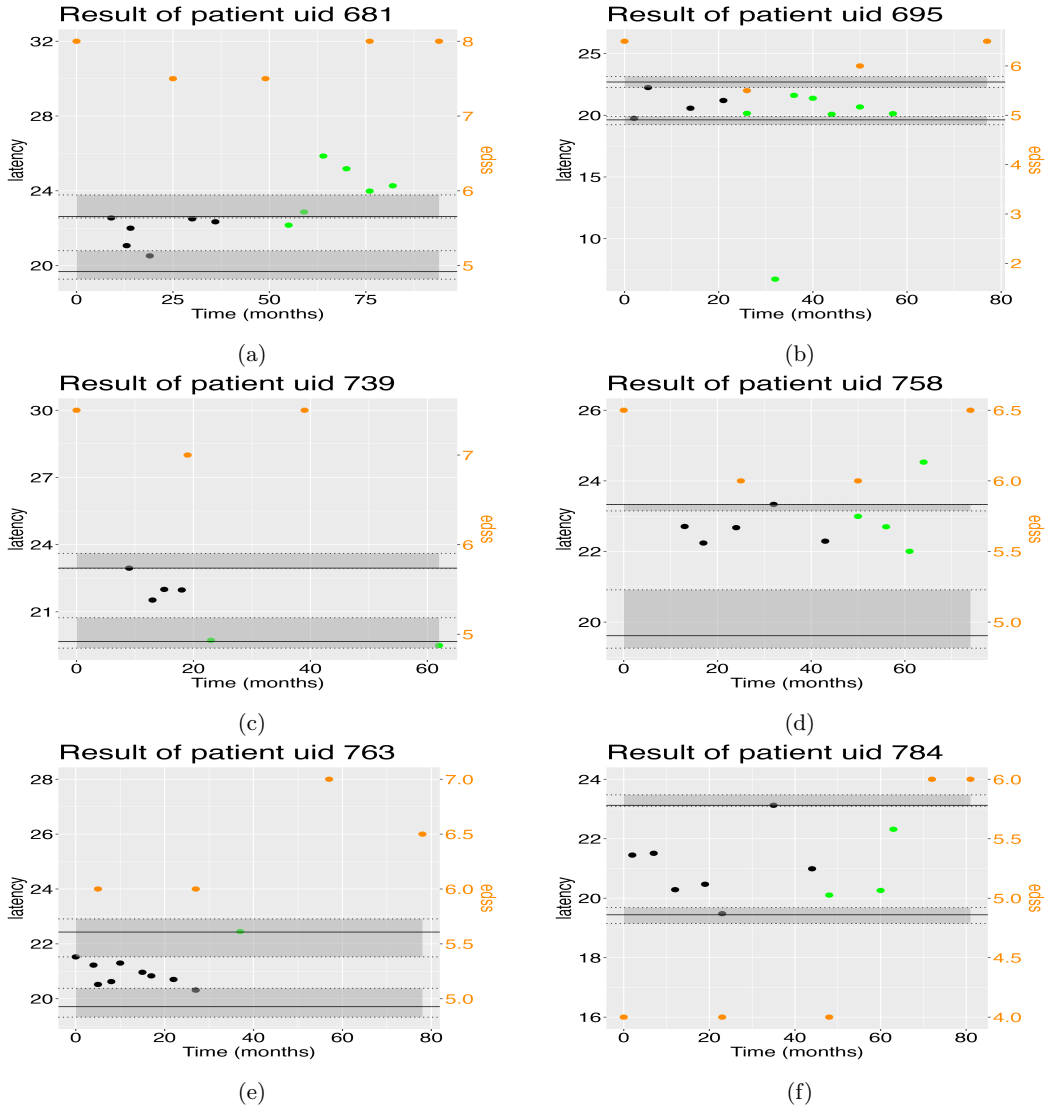


Figure 61: A subset of PwMS from the right hand data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression can be detected at 76 months based on EDSS score, and at 59 months based on IRI.
(b) Progression can be detected at 50 months based on EDSS score, and at 32 months based on IRI.
(c) Progression can be detected at 39 months based on EDSS score, and at 23 months based on IRI.
(d) Progression can be detected at 74 months based on EDSS score, and at 64 months based on IRI.
(e) Progression can be detected at 57 months based on EDSS score, and at 37 months based on IRI.
(f) Progression cannot be detected earlier by IRI than by EDSS score.

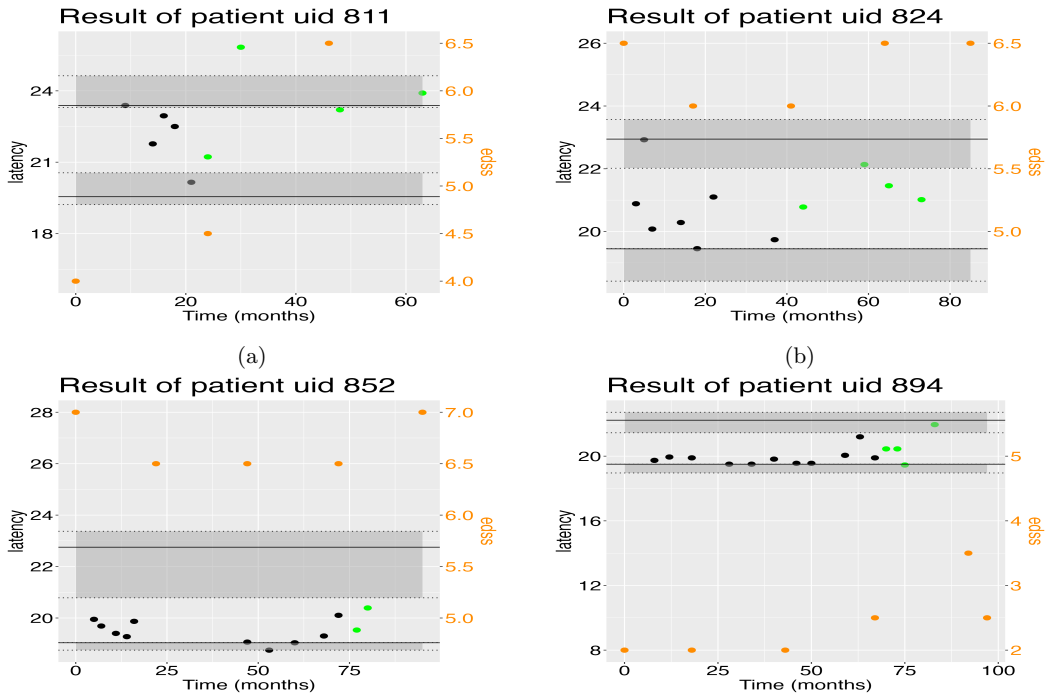
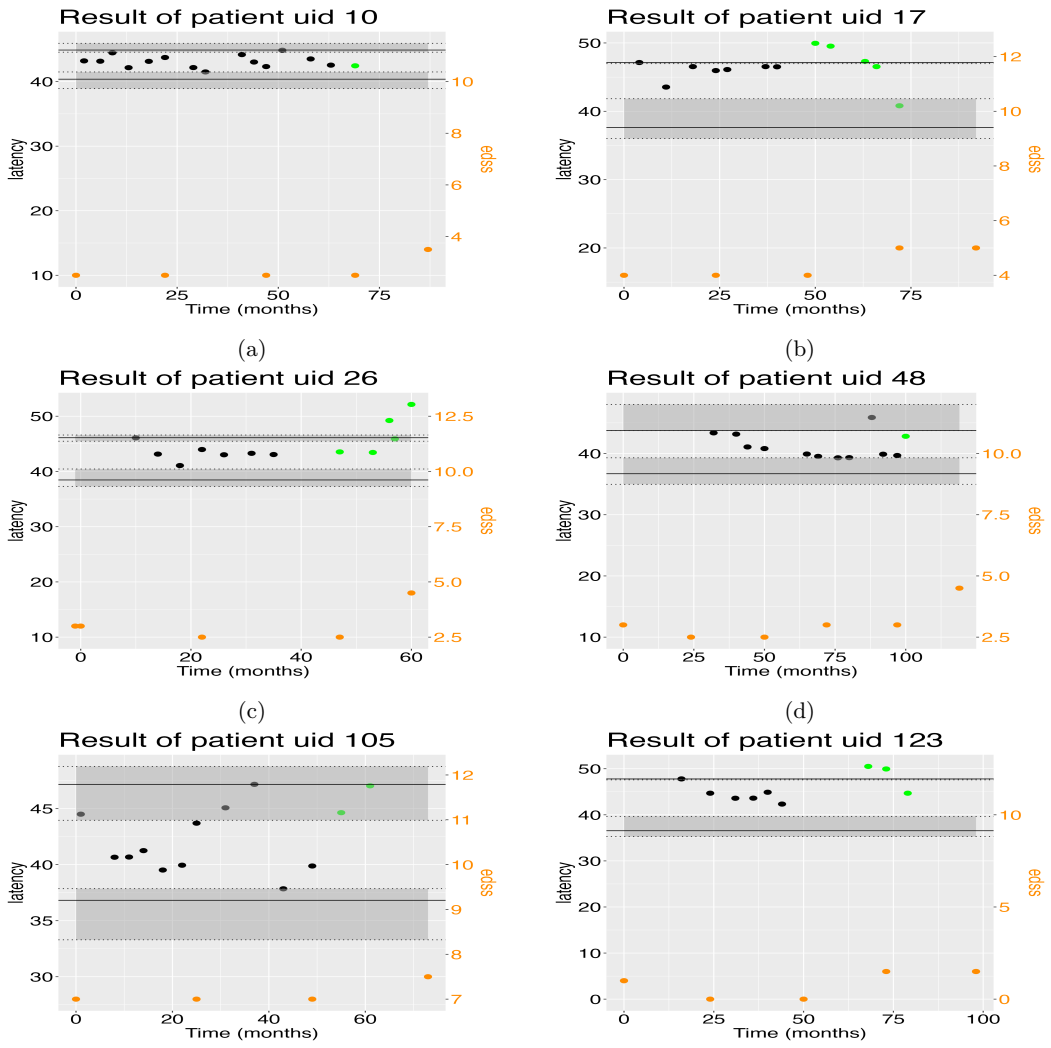


Figure 62: A subset of PwMS from the right hand data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression can be detected at 46 months based on EDSS score, and at 30 months based on IRI.
(b) Progression can be detected at 64 months based on EDSS score, and at 59 months based on IRI.
(c) Progression cannot be detected earlier by IRI than by EDSS score
(d) Progression can be detected at 92 months based on EDSS score, and at 75 months based on IRI.

I Confidence Interval of IRI and EDSS for right foot data



(a) (b) (c) (d) (e) (f)

Figure 63: A subset of PwMS from the right foot data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression cannot be detected earlier by IRI than by EDSS score due to no data during the time interval of 2 years.
(b) Progression can be detected at 72 months based on EDSS score, and at 50 months based on IRI.
(c) Progression can be detected at 60 months based on EDSS score, and at 56 months based on IRI.
(d) Progression cannot be detected earlier by IRI than by EDSS score
(e) Progression can be detected at 73 months based on EDSS score, and at 55 months based on IRI.
(f) Progression can be detected at 73 months based on EDSS score, and at 68 months based on IRI.

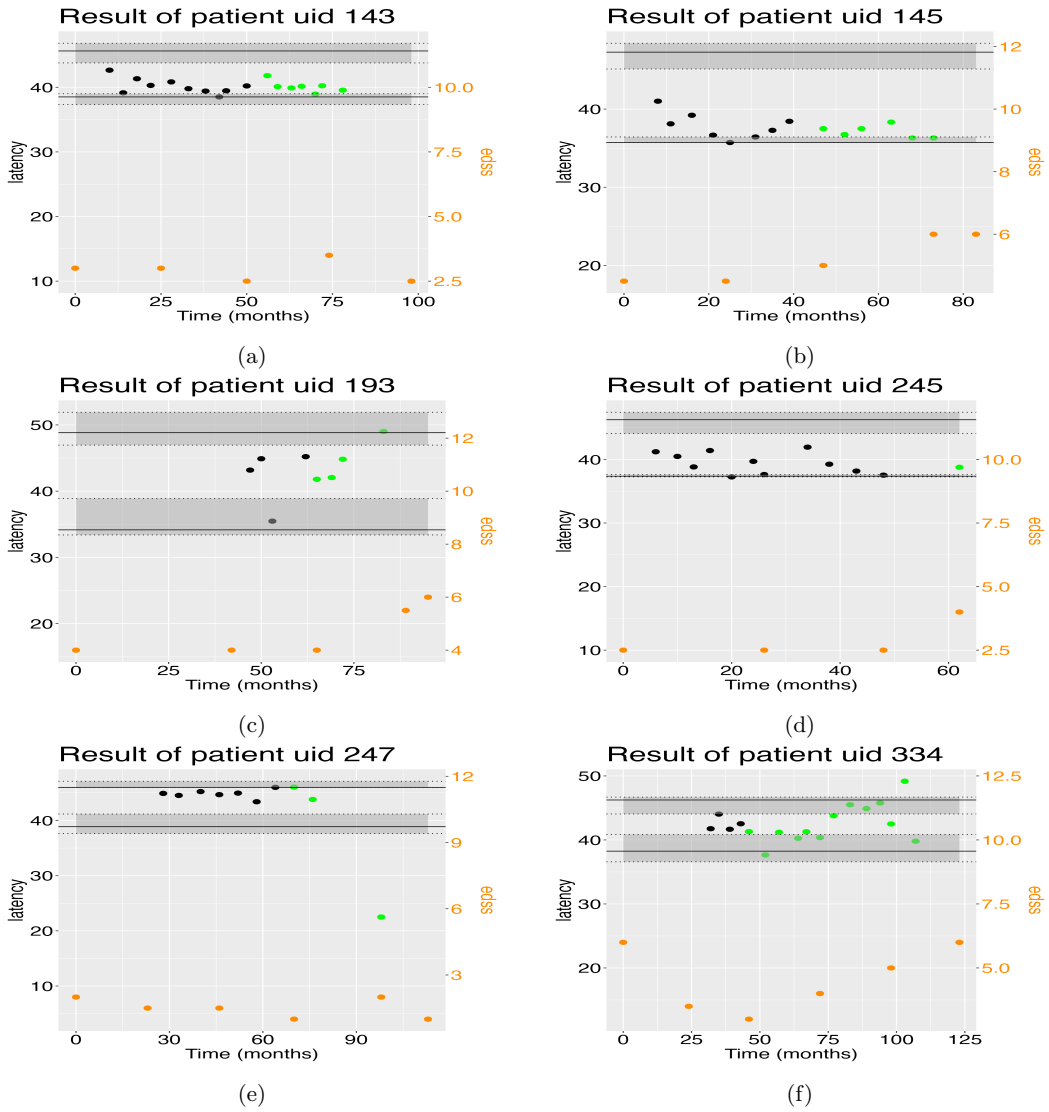


Figure 64: A subset of PwMS from the right foot data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression can be detected at 74 months based on EDSS score, and at 70 months based on IRI.
(b) Progression can be detected at 73 months based on EDSS score, and at 68 months based on IRI.
(c) Progression can be detected at 89 months based on EDSS score, and at 83 months based on IRI.
(d) Progression cannot be detected earlier by IRI than by EDSS score due to no data available during the time interval of 2 years.
(e) Progression can be detected at 98 months based on EDSS score, and at 70 months based on IRI.
(f) Progression can be detected at 72 months based on EDSS score, and at 52 months based on IRI.

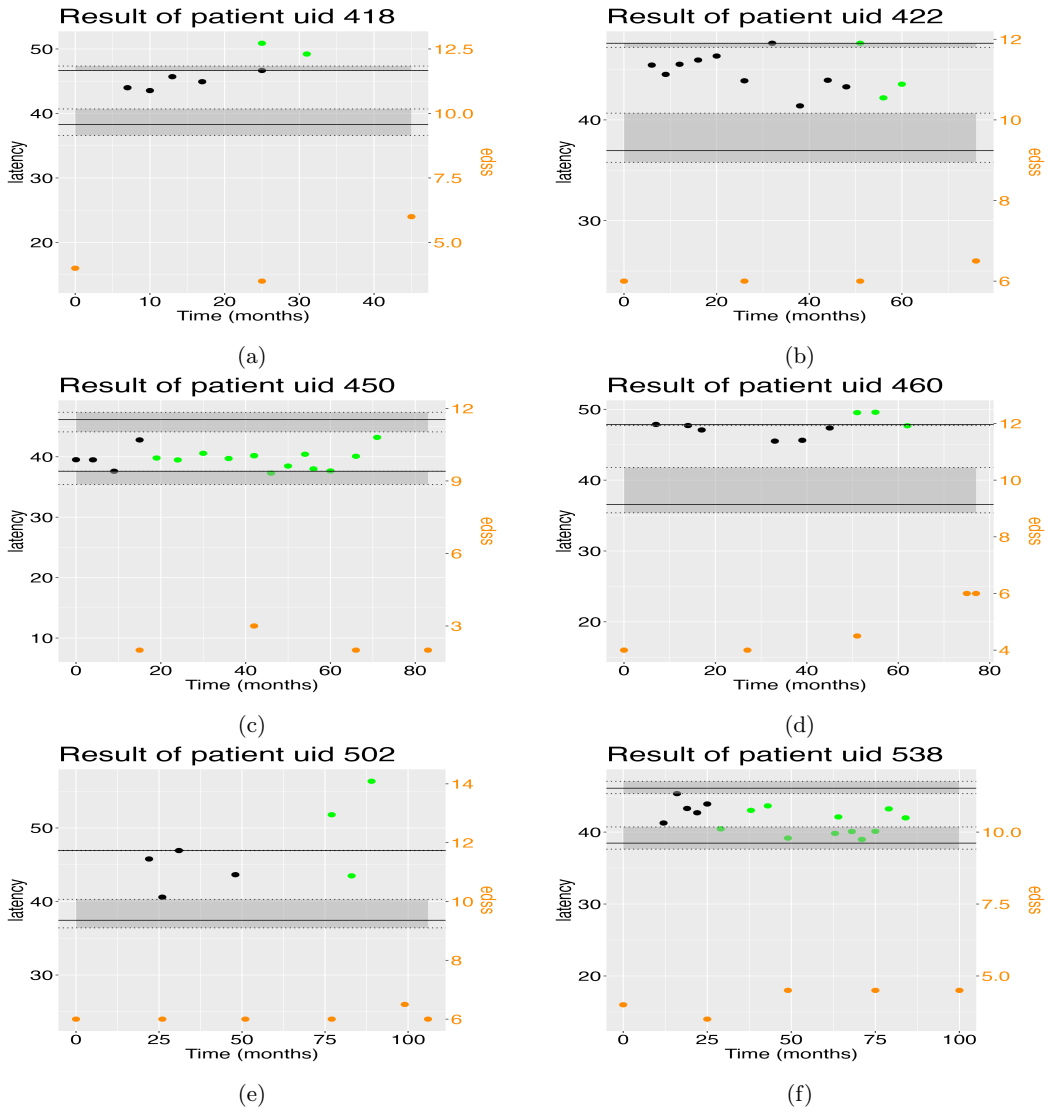


Figure 65: A subset of PwMS from the right foot data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression can be detected at 45 months based on EDSS score, and at 25 months based on IRI.
(b) Progression can be detected at 76 months based on EDSS score, and at 51 months based on IRI.
(c) Progression cannot be detected earlier by IRI than by EDSS score.
(d) Progression can be detected at 75 months based on EDSS score, and at 51 months based on IRI.
(e) Progression can be detected at 99 months based on EDSS score, and at 77 months based on IRI.
(f) Progression can be detected at 49 months based on EDSS score, and at 29 months based on IRI.

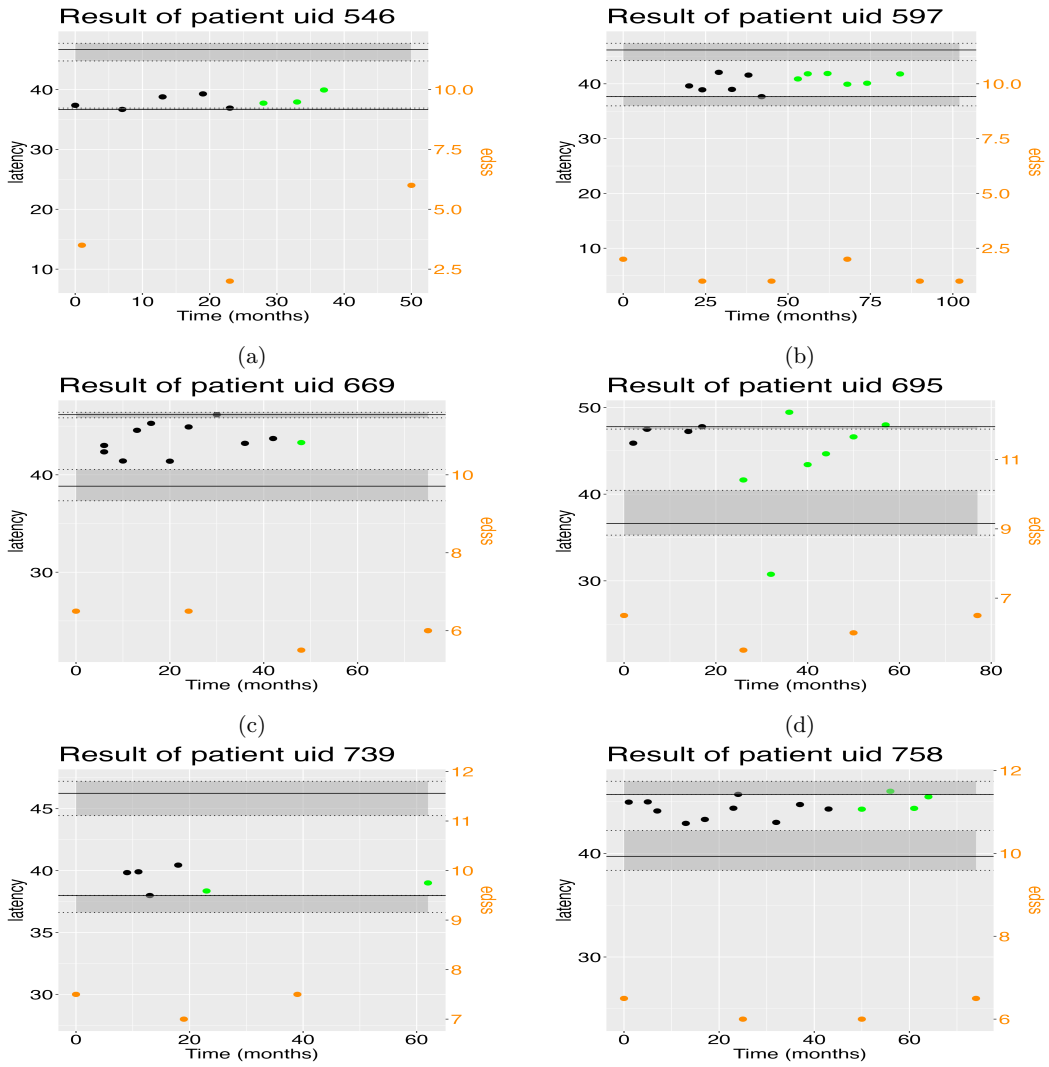


Figure 66: A subset of PwMS from the right foot data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression cannot be detected earlier by IRI than by EDSS score.
(b) Progression cannot be detected earlier by IRI than by EDSS score.
(c) Progression cannot be detected earlier by IRI than by EDSS score due to no data available during the time interval of 2 years.
(d) Progression can be detected at 50 months based on EDSS score, and at 32 months based on IRI.
(e) Progression cannot be detected earlier by IRI than by EDSS score.
(f) Progression can be detected at 74 months based on EDSS score, and at 56 months based on IRI.

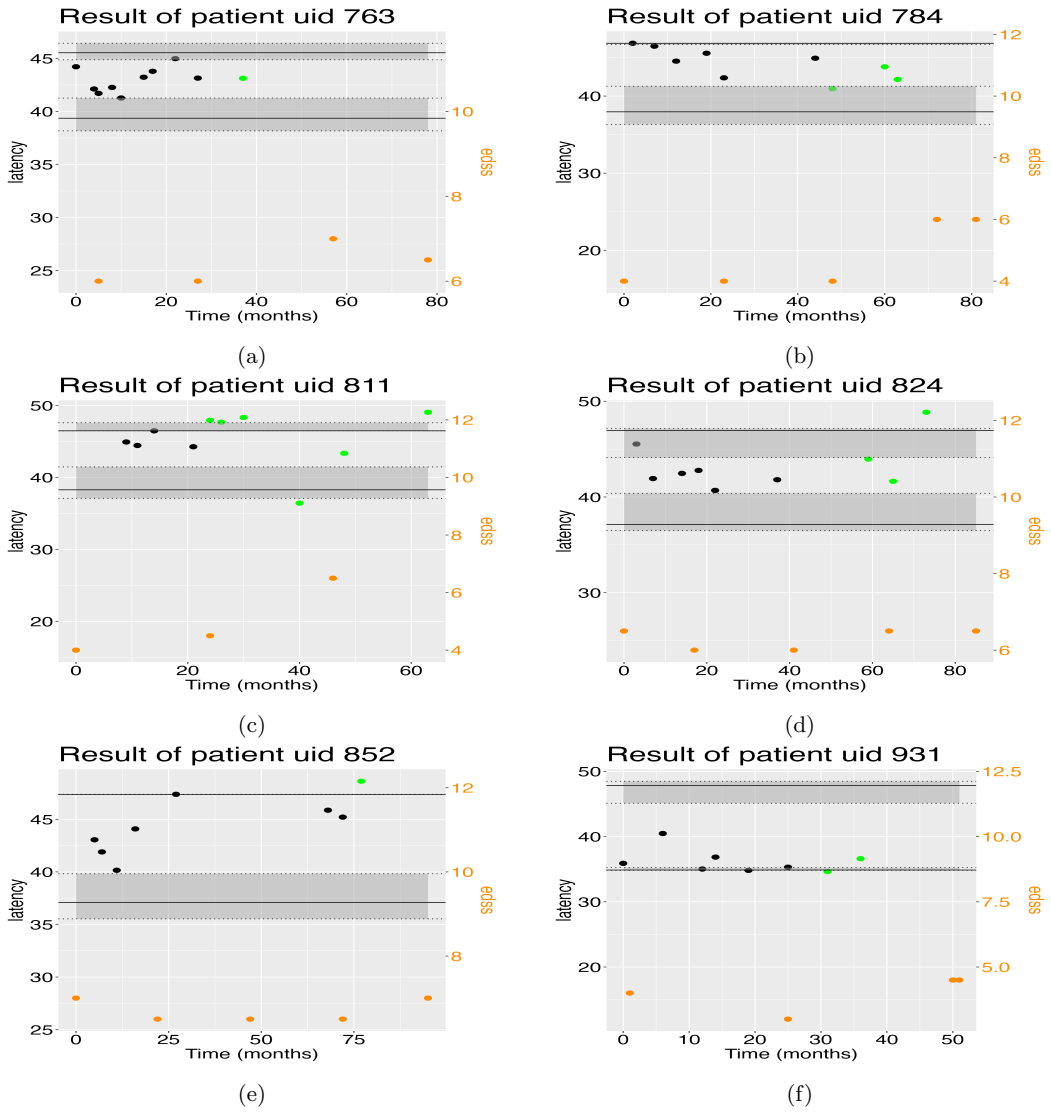
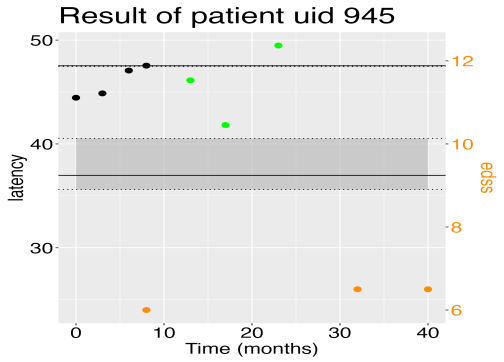


Figure 67: A subset of PwMS from the right foot data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression cannot be detected earlier by IRI than by EDSS score.
(b) Progression can be detected at 72 months based on EDSS score, and at 48 months based on IRI.
(c) Progression can be detected at 46 months based on EDSS score, and at 24 months based on IRI.
(d) Progression can be detected at 64 months based on EDSS score, and at 59 months based on IRI.
(e) Progression can be detected at 95 months based on EDSS score, and at 77 months based on IRI.
(f) Progression can be detected at 50 months based on EDSS score, and at 31 months based on IRI.



(a)

Figure 68: A subset of PwMS from the right foot data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points present the EDSS score. The progression of MS can be detected earlier using IRI than EDSS score. (a) Progression can be detected at 32 months based on EDSS score, and at 23 months based on IRI.

J Confidence Interval of IRI and EDSS for left hand data

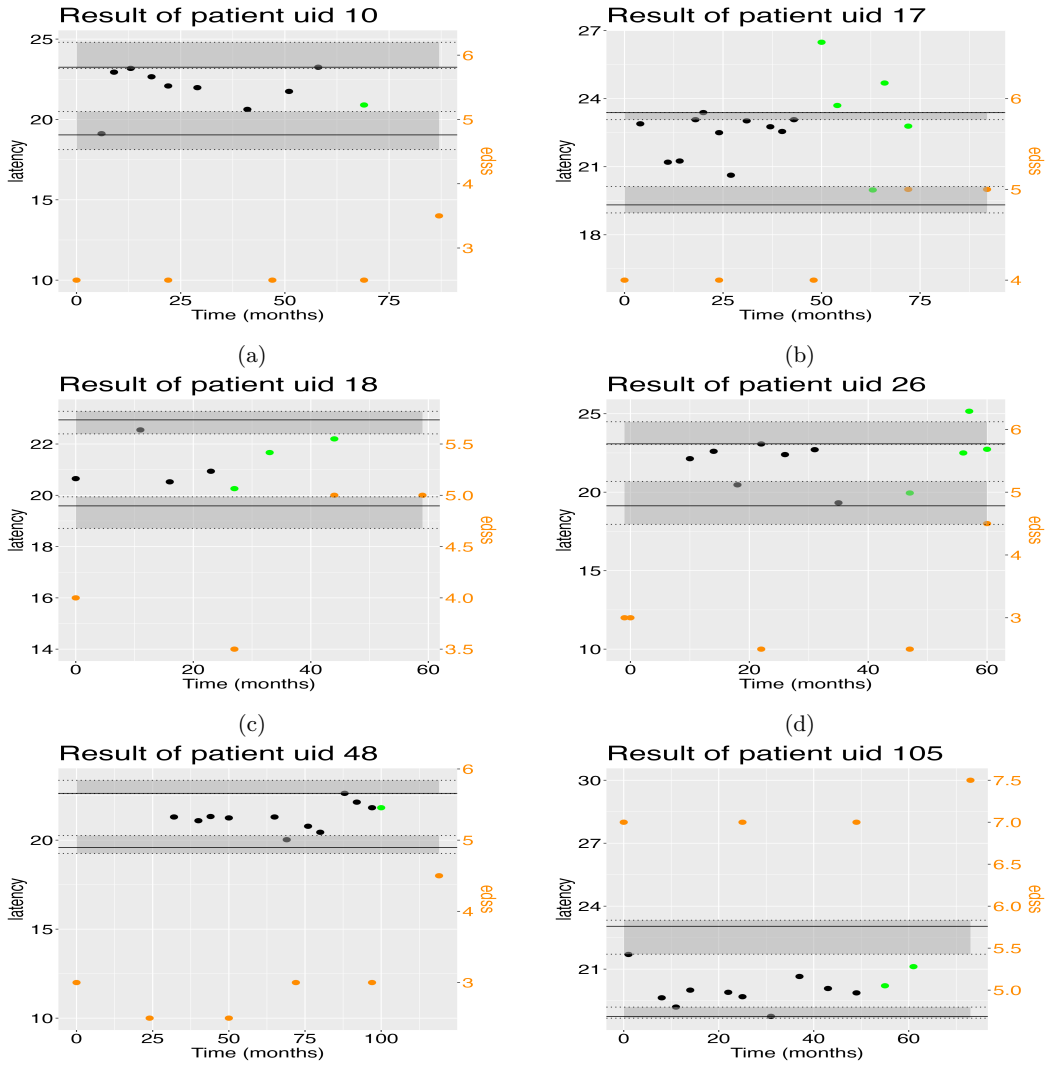


Figure 69: A subset of PwMS from the left hand data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression cannot be detected earlier by IRI than by EDSS score.
(b) Progression can be detected at 72 months based on EDSS score, and at 50 months based on IRI.
(c) Progression cannot be detected earlier by IRI than by EDSS score.
(d) Progression can be detected at 60 months based on EDSS score, and at 47 months based on IRI.
(e) Progression cannot be detected earlier by IRI than by EDSS score.
(f) Progression cannot be detected earlier by IRI than by EDSS score.

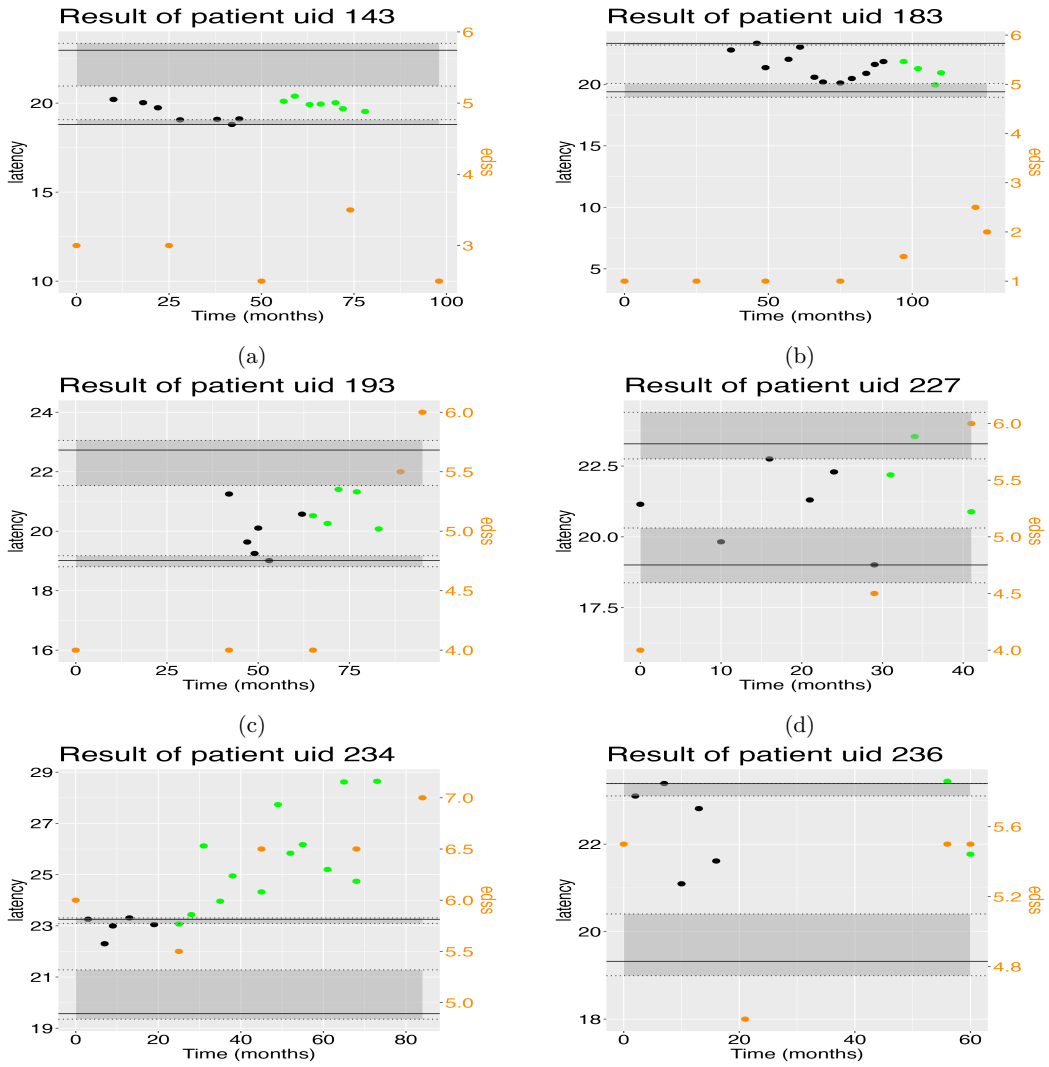


Figure 70: A subset of PwMS from the left hand data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points present the EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression cannot be detected earlier by IRI than by EDSS score.
(b) Progression can be detected at 122 months based on EDSS score, and at 108 months based on IRI.
(c) Progression cannot be detected earlier by IRI than by EDSS score.
(d) Progression can be detected at 41 months based on EDSS score, and at 34 months based on IRI.
(e) Progression can be detected at 45 months based on EDSS score, and at 25 months based on IRI.
(f) Progression cannot be detected earlier by IRI than by EDSS score due to no data available during the time interval of 2 years.

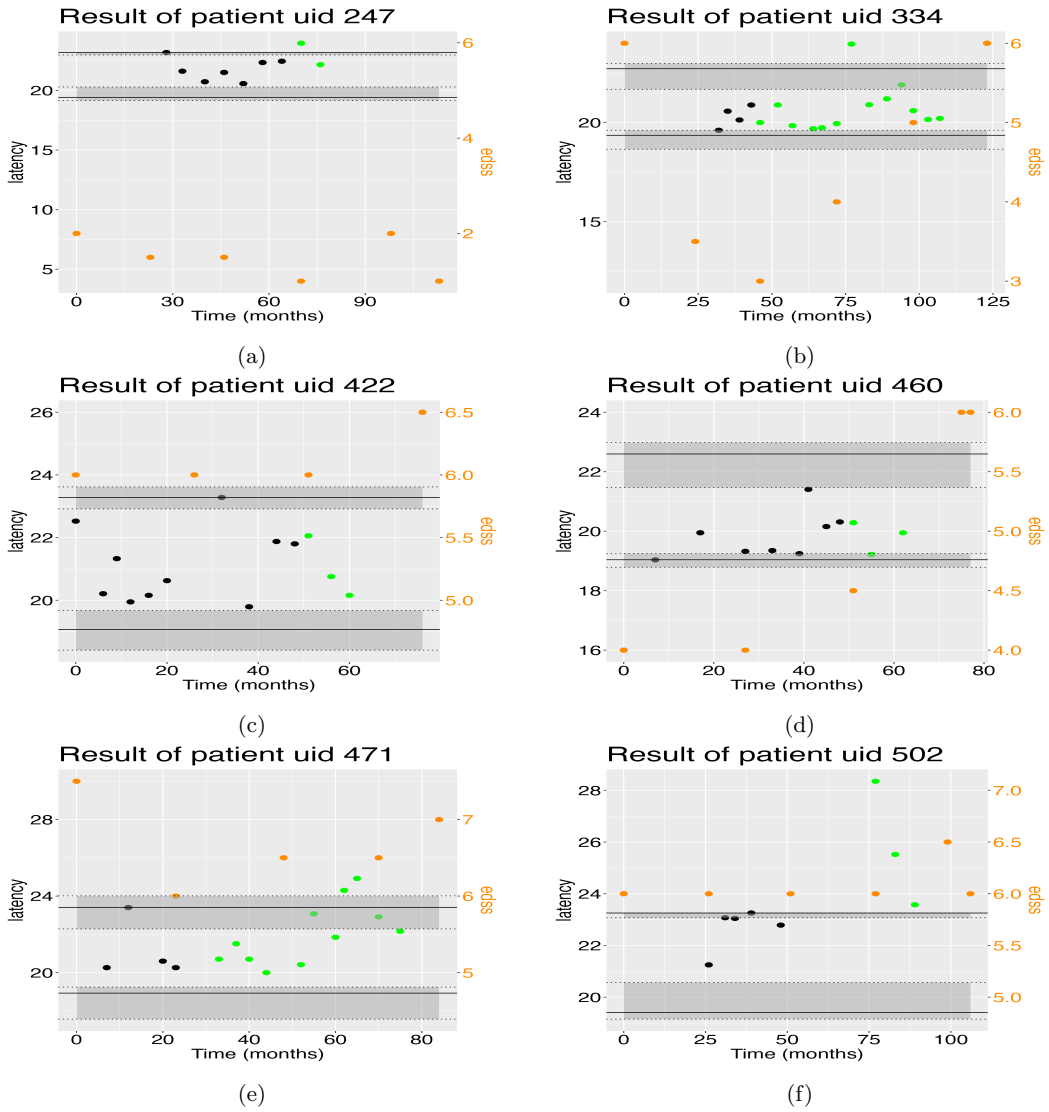


Figure 71: A subset of PwMS from the left hand data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression can be detected at 98 months based on EDSS score, and at 70 months based on IRI.
(b) Progression cannot be detected earlier by IRI than by EDSS score.
(c) Progression cannot be detected earlier by IRI than by EDSS score.
(d) Progression can be detected at 75 months based on EDSS score, and at 55 months based on IRI.
(e) Progression cannot be detected earlier by IRI than by EDSS score.
(f) Progression can be detected at 99 months based on EDSS score, and at 77 months based on IRI.

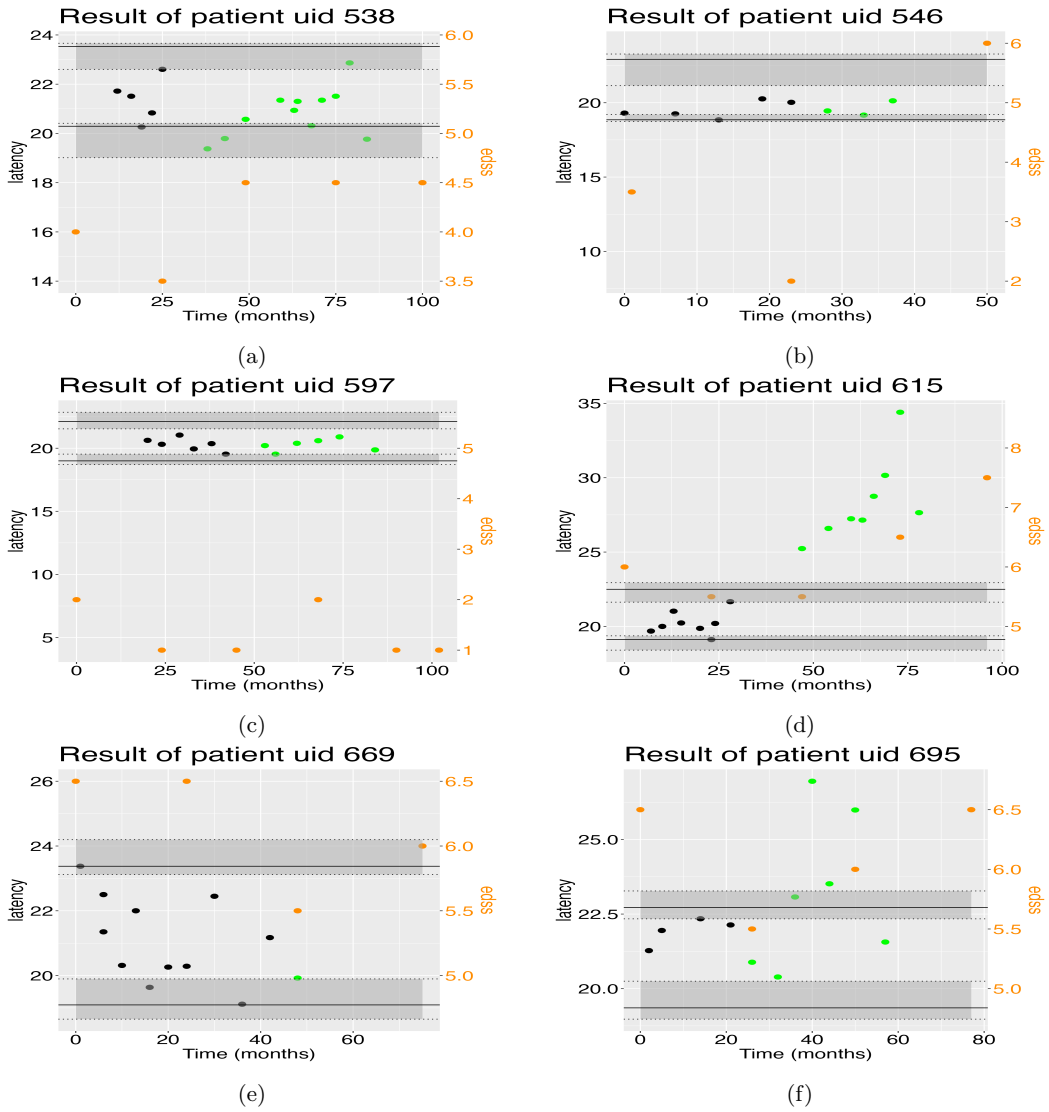


Figure 72: A subset of PwMS from the left hand data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points present the EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression can be detected at 49 months based on EDSS score, and at 38 months based on IRI.
(b) Progression can be detected at 50 months based on EDSS score, and at 33 months based on IRI.
(c) Progression can be detected at 68 months based on EDSS score, and at 56 months based on IRI.
(d) Progression can be detected at 73 months based on EDSS score, and at 47 months based on IRI.
(e) Progression cannot be detected earlier by IRI than by EDSS score.
(f) Progression can be detected at 50 months based on EDSS score, and at 36 months based on IRI.

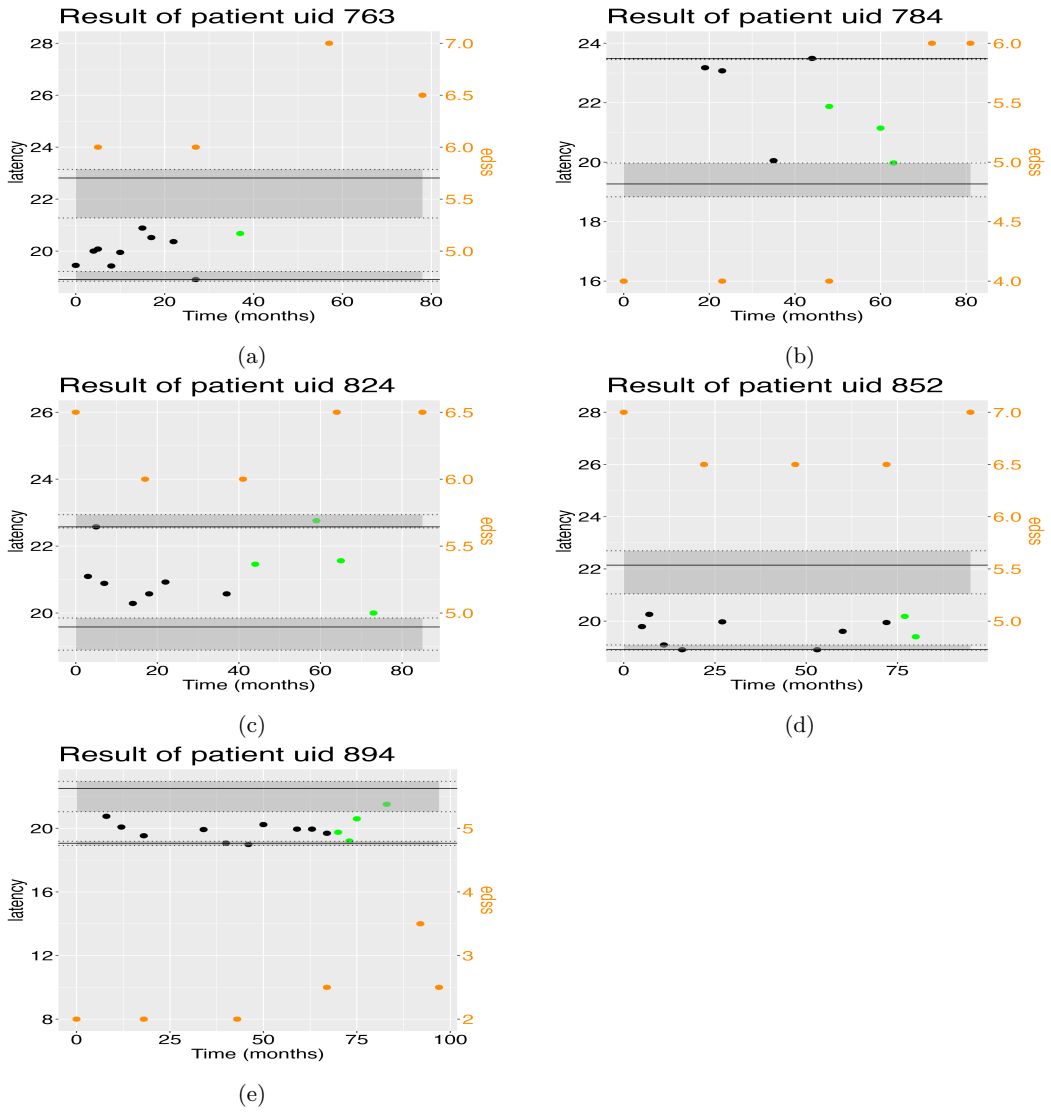


Figure 73: A subset of PwMS from the left hand data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression cannot be detected earlier by IRI than by EDSS score.
(b) Progression can be detected at 72 months based on EDSS score, and at 63 months based on IRI.
(c) Progression can be detected at 64 months based on EDSS score, and at 59 months based on IRI.
(d) Progression cannot be detected earlier by IRI than by EDSS score.
(e) Progression can be detected at 76 months based on EDSS score, and at 59 months based on IRI.
(f) Progression can be detected at 92 months based on EDSS score, and at 73 months based on IRI.

K Confidence Interval of IRI and EDSS left foot data

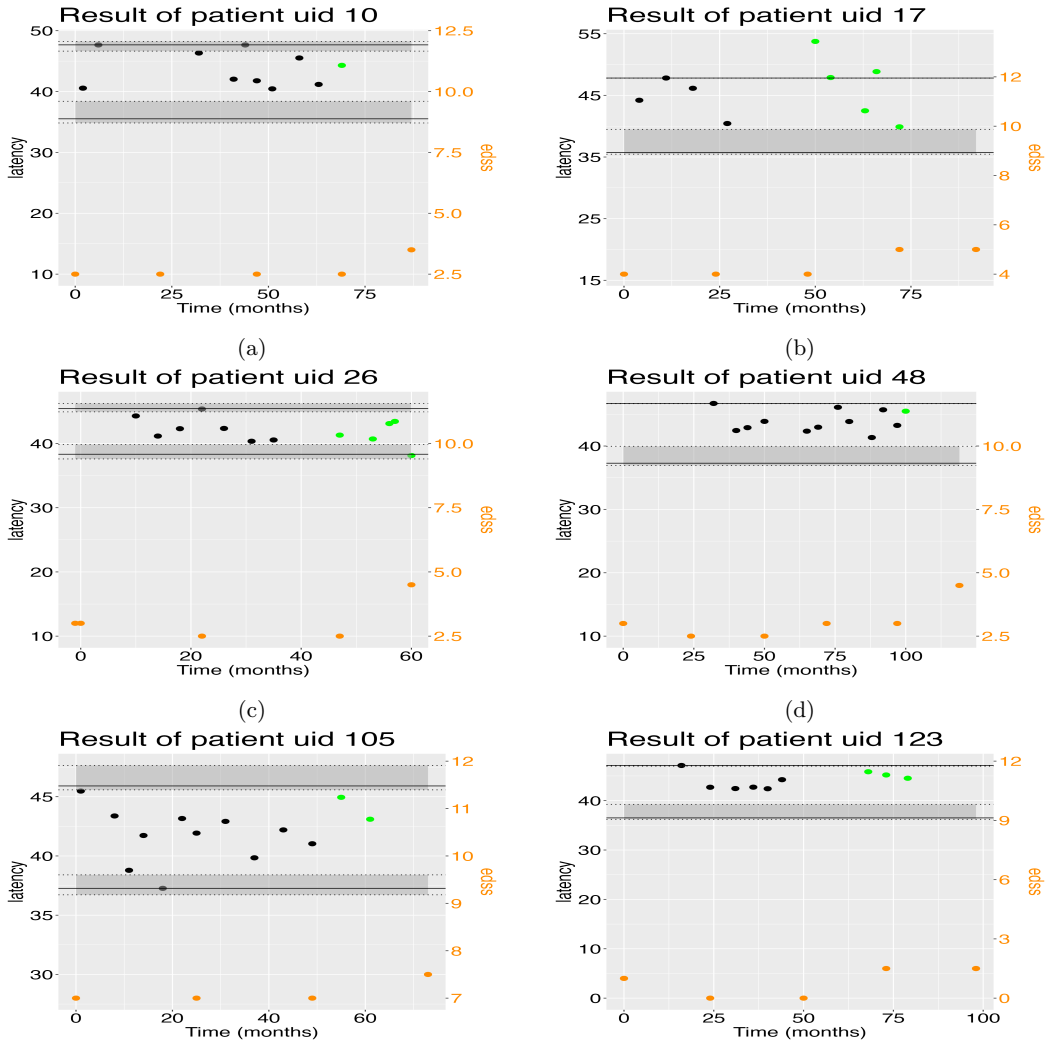


Figure 74: A subset of PwMS from the left foot data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points present the EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression cannot be detected earlier by IRI than by EDSS score.
(b) Progression can be detected at 72 months based on EDSS score, and at 50 months based on IRI.
(c) Progression cannot be detected earlier by IRI than by EDSS score.
(d) Progression cannot be detected earlier by IRI than by EDSS score.
(e) Progression cannot be detected earlier by IRI than by EDSS score.
(f) Progression cannot be detected earlier by IRI than by EDSS score.

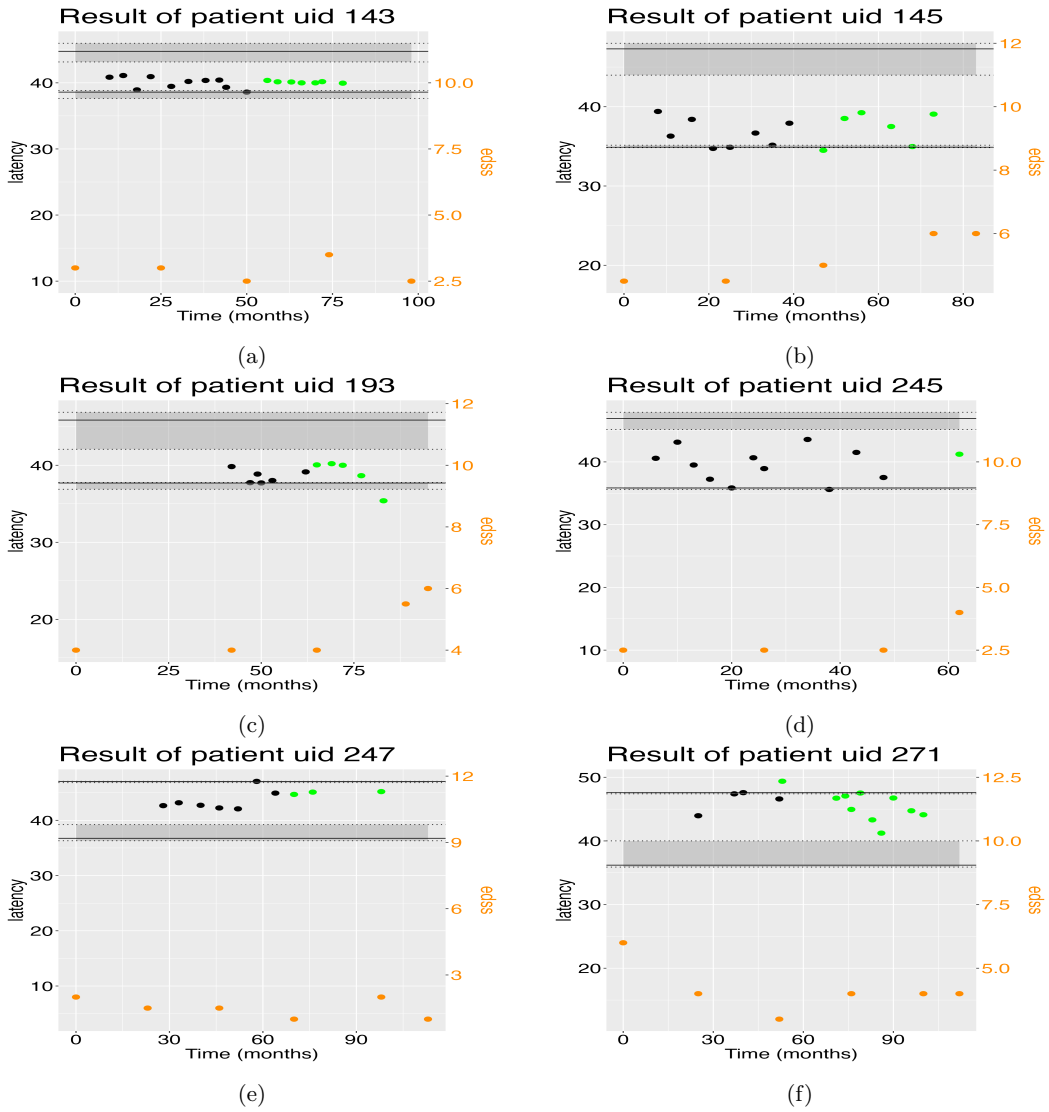


Figure 75: A subset of PwMS from the left foot data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression cannot be detected earlier by IRI than by EDSS score.
(b) Progression can be detected at 73 months based on EDSS score, and at 47 months based on IRI.
(c) Progression can be detected at 89 months based on EDSS score, and at 83 months based on IRI.
(d)Progression cannot be detected earlier by IRI than by EDSS score.
(e) Progression cannot be detected earlier by IRI than by EDSS score.
(f) Progression can be detected at 76 months based on EDSS score, and at 53 months based on IRI.

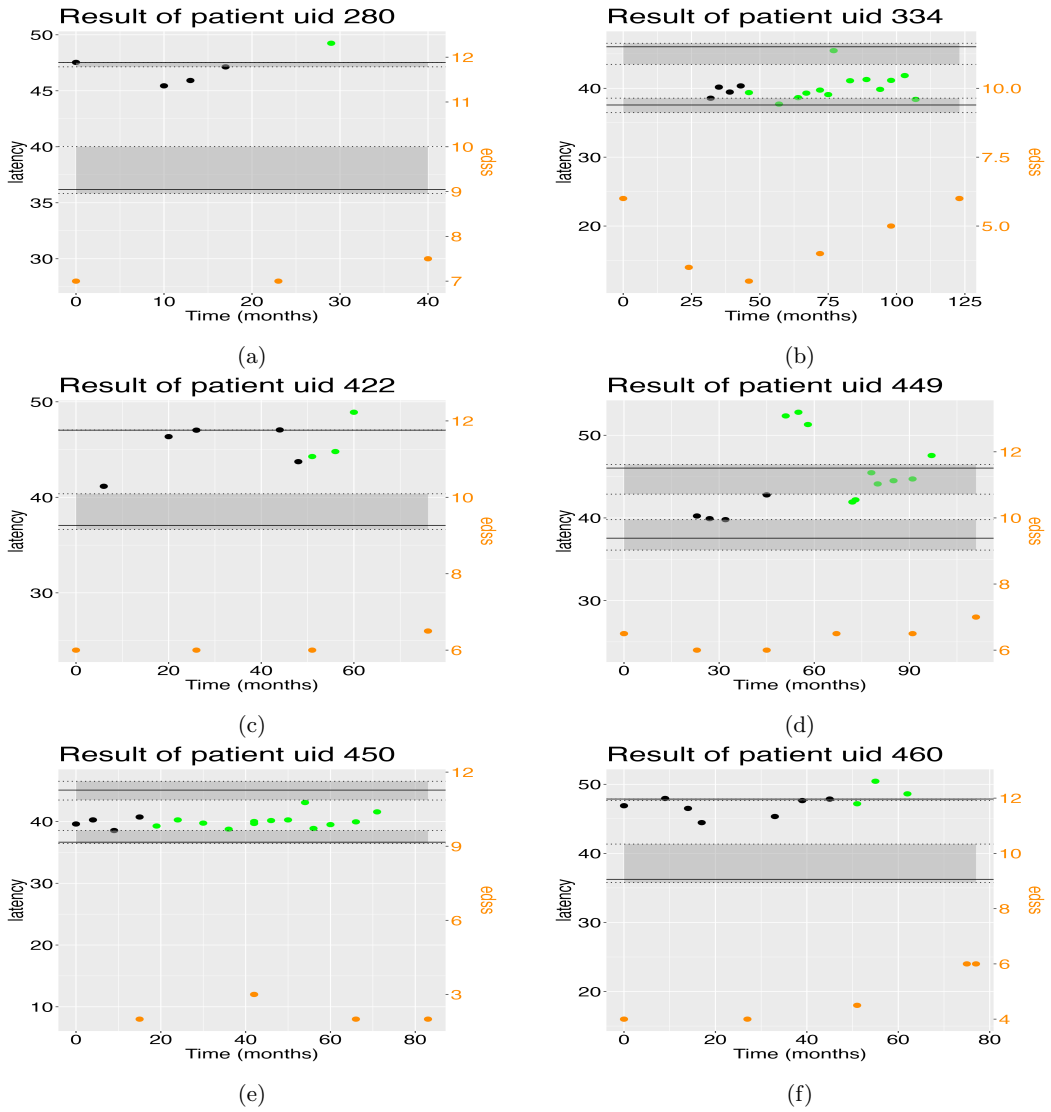


Figure 76: A subset of PwMS from the left foot data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression can be detected at 40 months based on EDSS score, and at 29 months based on IRI.
(b) Progression can be detected at 72 months based on EDSS score, and at 57 months based on IRI.
(c) Progression can be detected at 76 months based on EDSS score, and at 60 months based on IRI.
(d) Progression can be detected at 67 months based on EDSS score, and at 251 months based on IRI.
(e)Progression cannot be detected earlier by IRI than by EDSS score.
(f) Progression can be detected at 75 months based on EDSS score, and at 55 months based on IRI.

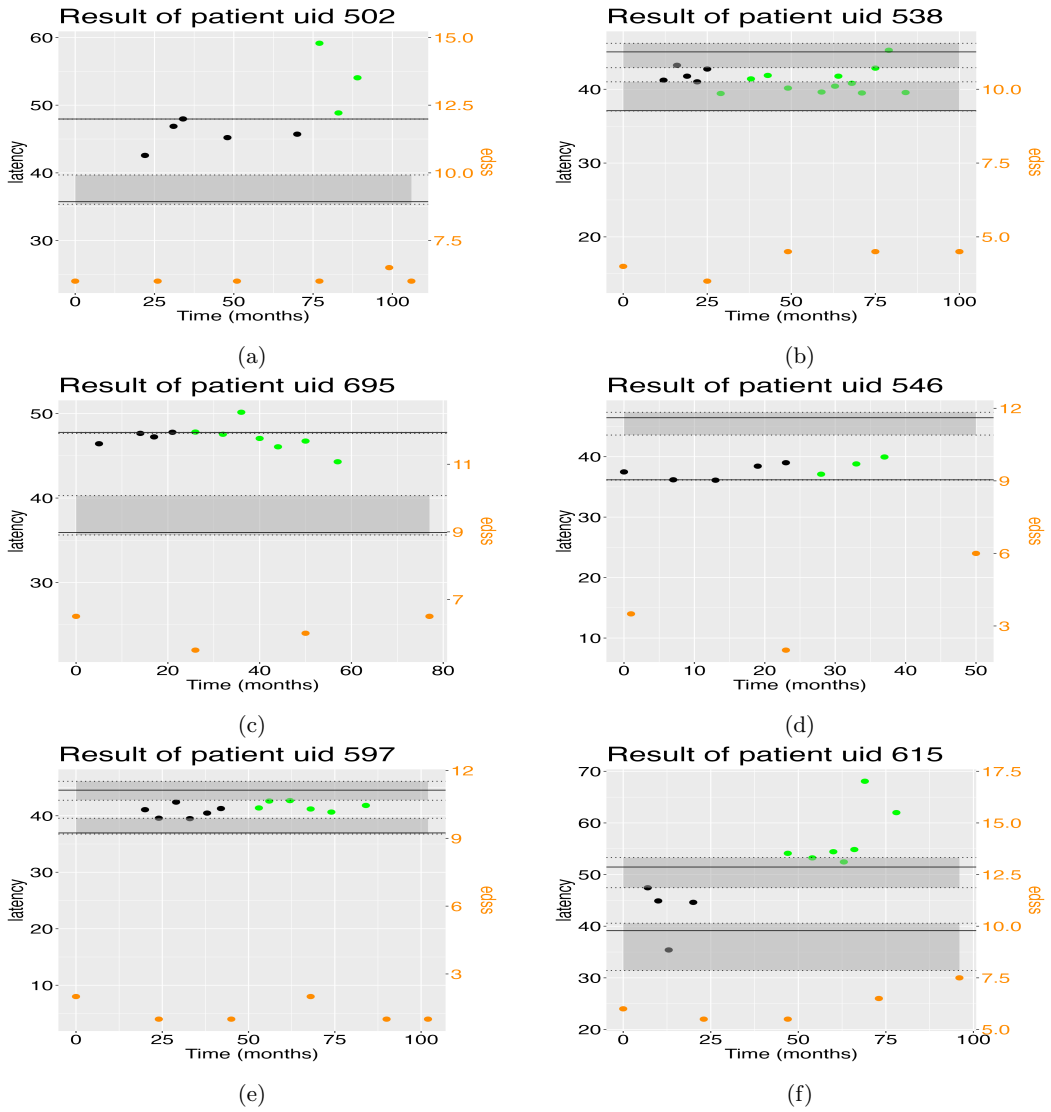


Figure 77: A subset of PwMS from the left foot data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points present the EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression can be detected at 99 months based on EDSS score, and at 77 months based on IRI.
(b) Progression can be detected at 49 months based on EDSS score, and at 29 months based on IRI.
(c) Progression can be detected at 50 months based on EDSS score, and at 26months based on IRI.
(d) Progression cannot be detected earlier by IRI than by EDSS score.
(e) Progression can be detected at 68 months based on EDSS score, and at 56 months based on IRI.
(f) Progression can be detected at 73 months based on EDSS score, and at 47 months based on IRI.

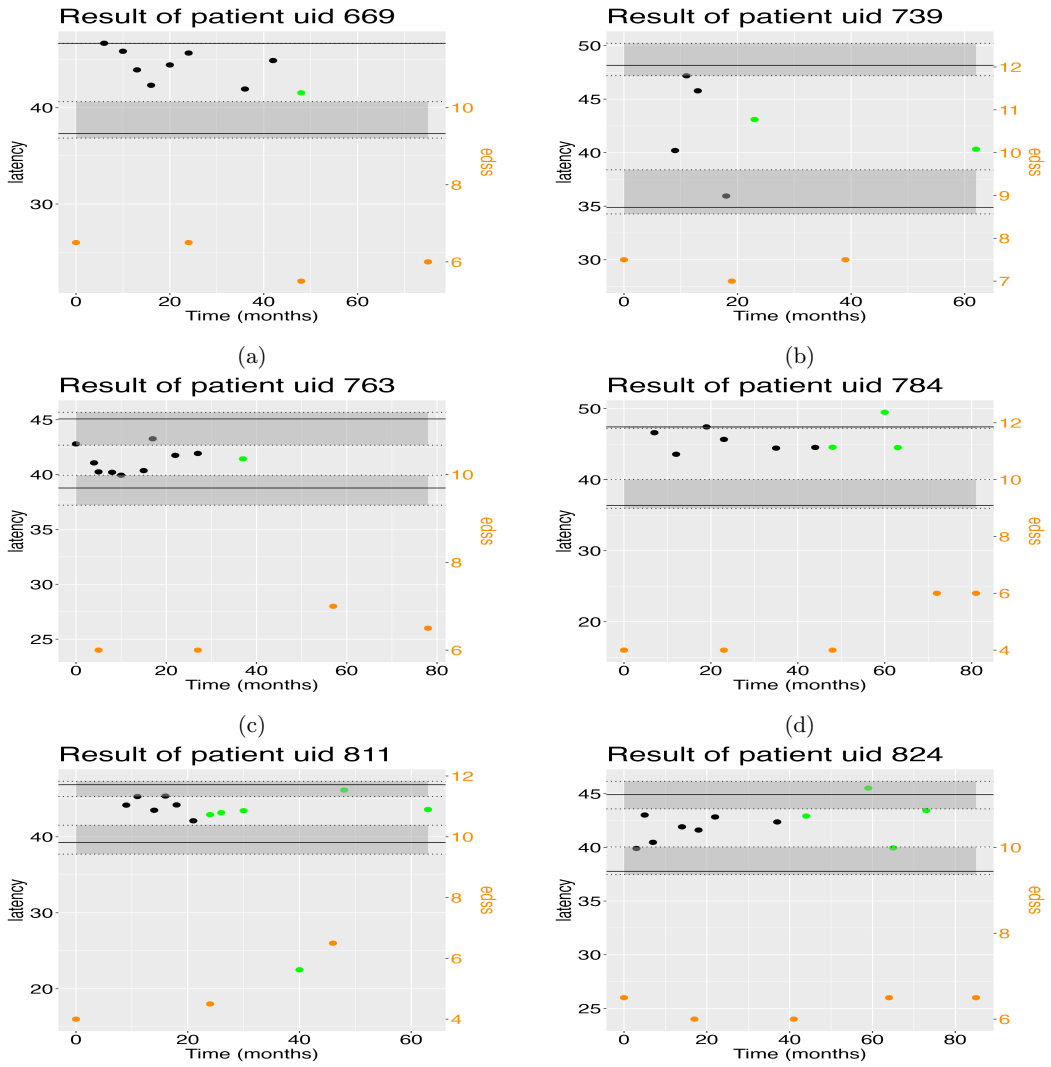


Figure 78: A subset of PwMS from the left foot data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression cannot be detected earlier by IRI than by EDSS score.
(b)Progression cannot be detected earlier by IRI than by EDSS score.
(c) Progression cannot be detected earlier by IRI than by EDSS score.
(d) Progression can be detected at 72 months based on EDSS score, and at 60 months based on IRI.
(e) Progression can be detected at 46 months based on EDSS score, and at 40 months based on IRI.
(f) Progression can be detected at 64 months based on EDSS score, and at 59 months based on IRI.

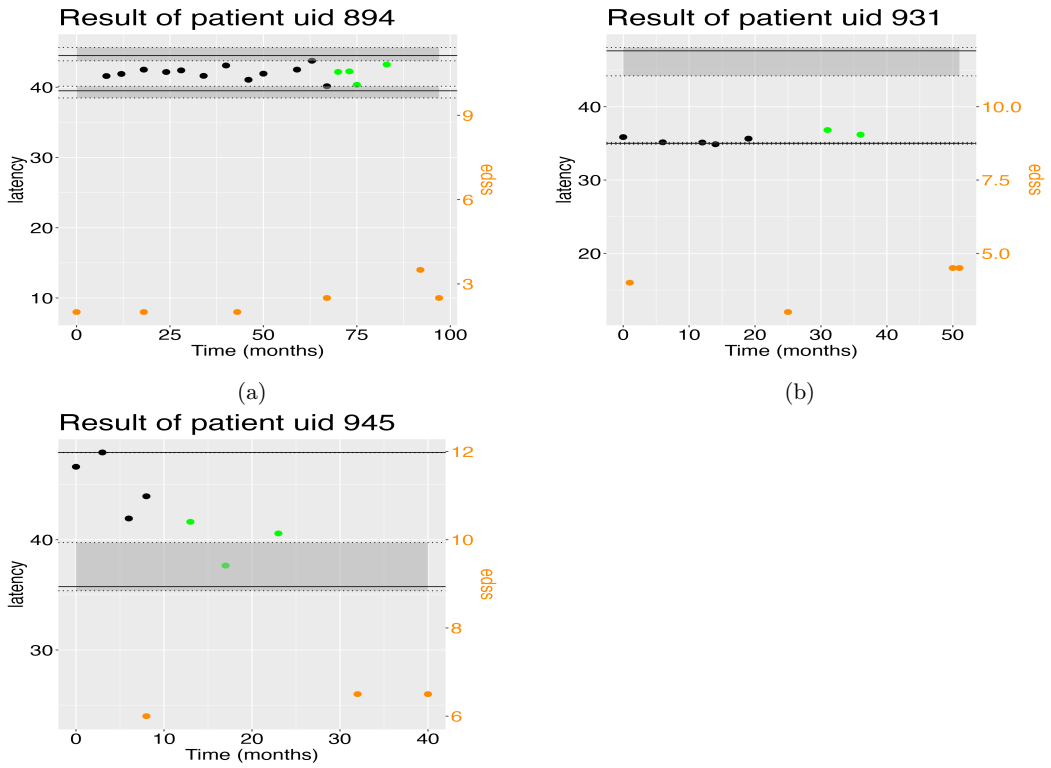


Figure 79: A subset of PwMS from the left foot data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points present the EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression cannot be detected earlier by IRI than by EDSS score.
(b) Progression cannot be detected earlier by IRI than by EDSS score.
(c) Progression can be detected at 32 months based on EDSS score, and at 17 months based on IRI.

L Confidence Interval of IRI and full EDSS for right hand data

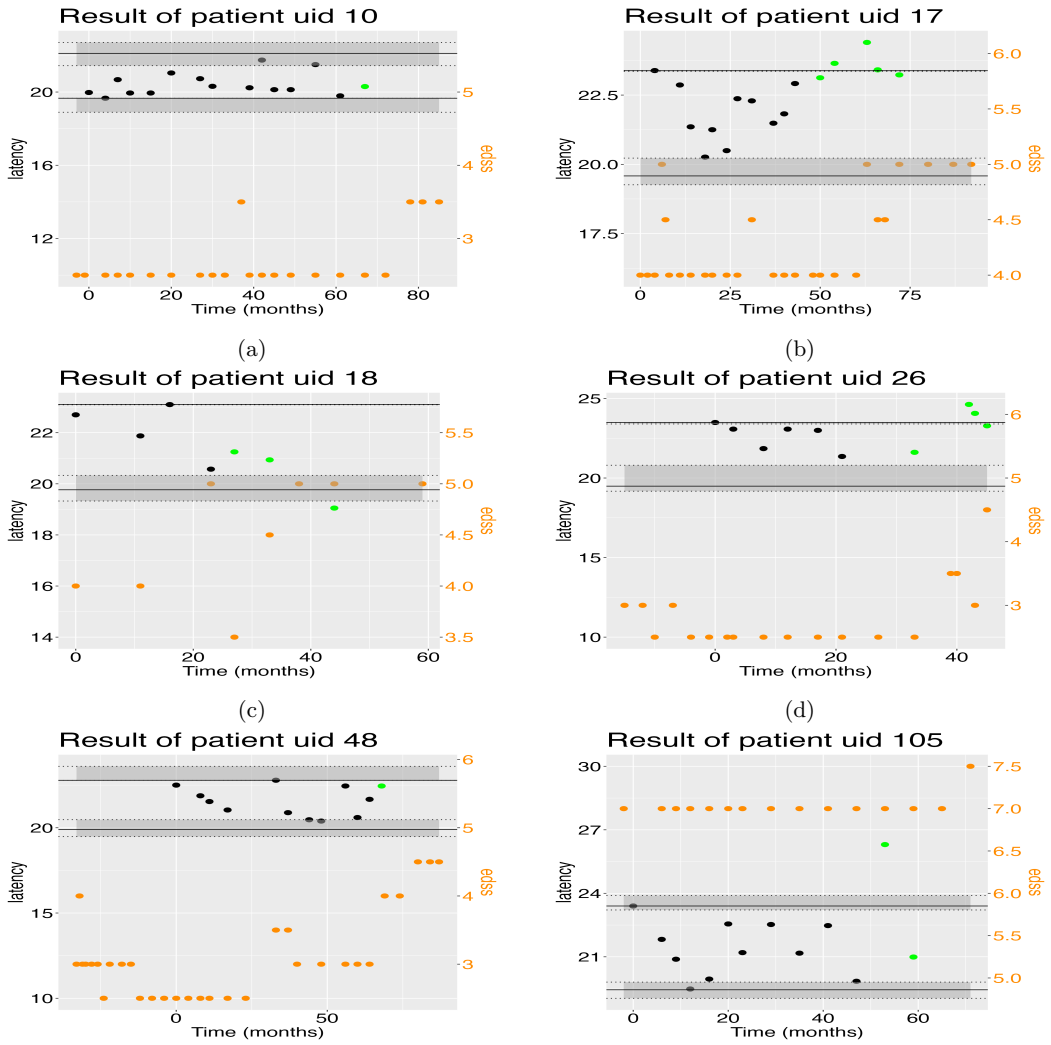


Figure 80: A subset of PwMS from the right hand data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points present the EDSS score. The progression of MS can be detected earlier using IRI on latency measurement than EDSS score.
(a) No data available to detect the progression based on IRI
(b) Progression can be detected at 72 months based on EDSS score, and at 54 months based on IRI.
(c) Progression cannot be detected sooner by IRI than by EDSS score
(d) Progression can be detected at 60 months based on EDSS score, and at 56 months based on IRI.
(e) No data available to detect the progression based on IRI.
(f) Progression can be detected at 73 months based on EDSS score, and at 55 months based on IRI.

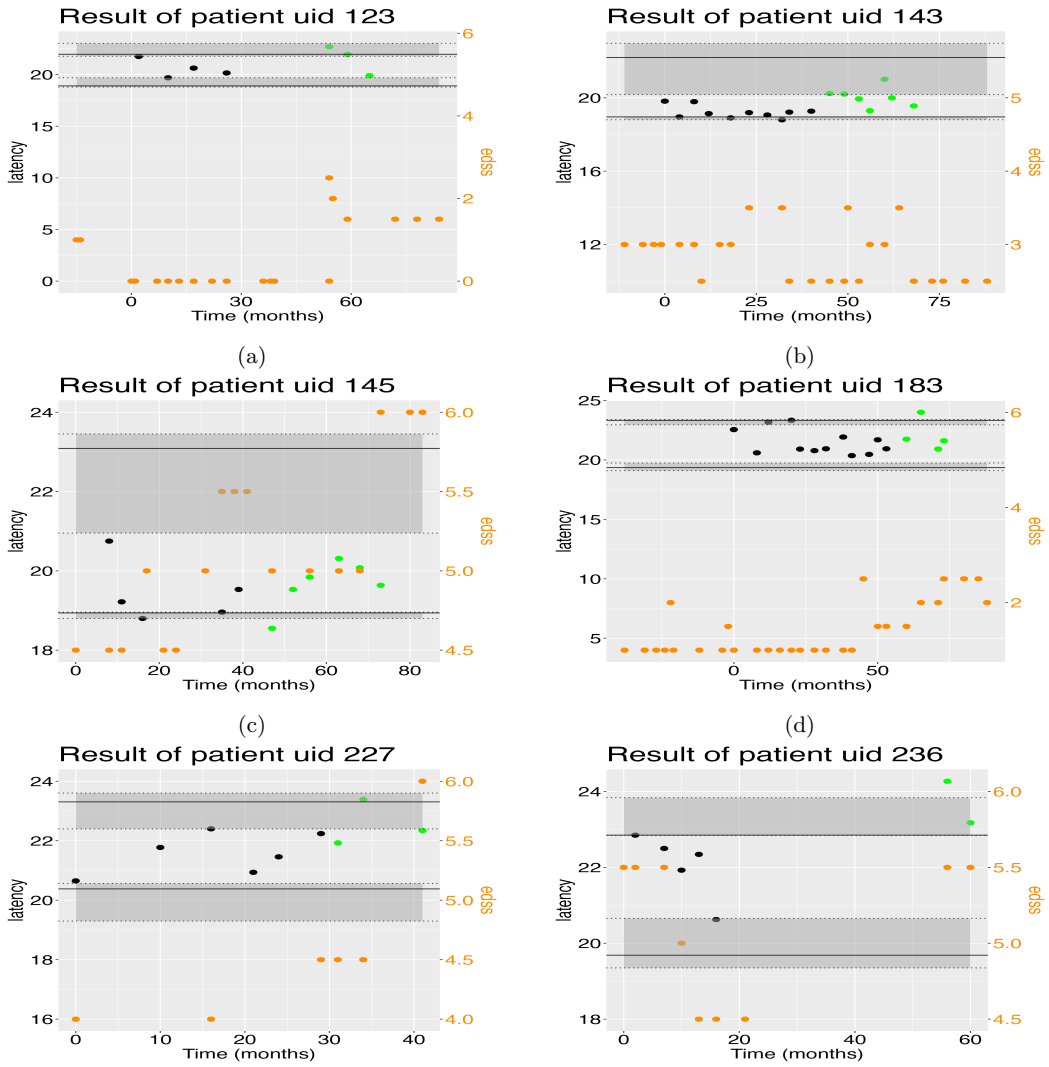


Figure 81: A subset of PwMS from the right hand data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points present the EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression can be detected at 73 months based on EDSS score, and at 68 months based on IRI.
(b) Progression can be detected at 74 months based on EDSS score, and at 56 months based on IRI.
(c) Progression cannot be detected earlier by IRI than EDSS score due to no data vailable during the time interval of 2 years
(d) Progression can be detected at 122 months based on EDSS score, and at 102 months based on IRI.
(e) Progression can be detected at 41 months based on EDSS score, and at 34 months based on IRI.
(f) Progression cannot be detected earlier by IRI than EDSS score due to no data available during the time interval of 2 years

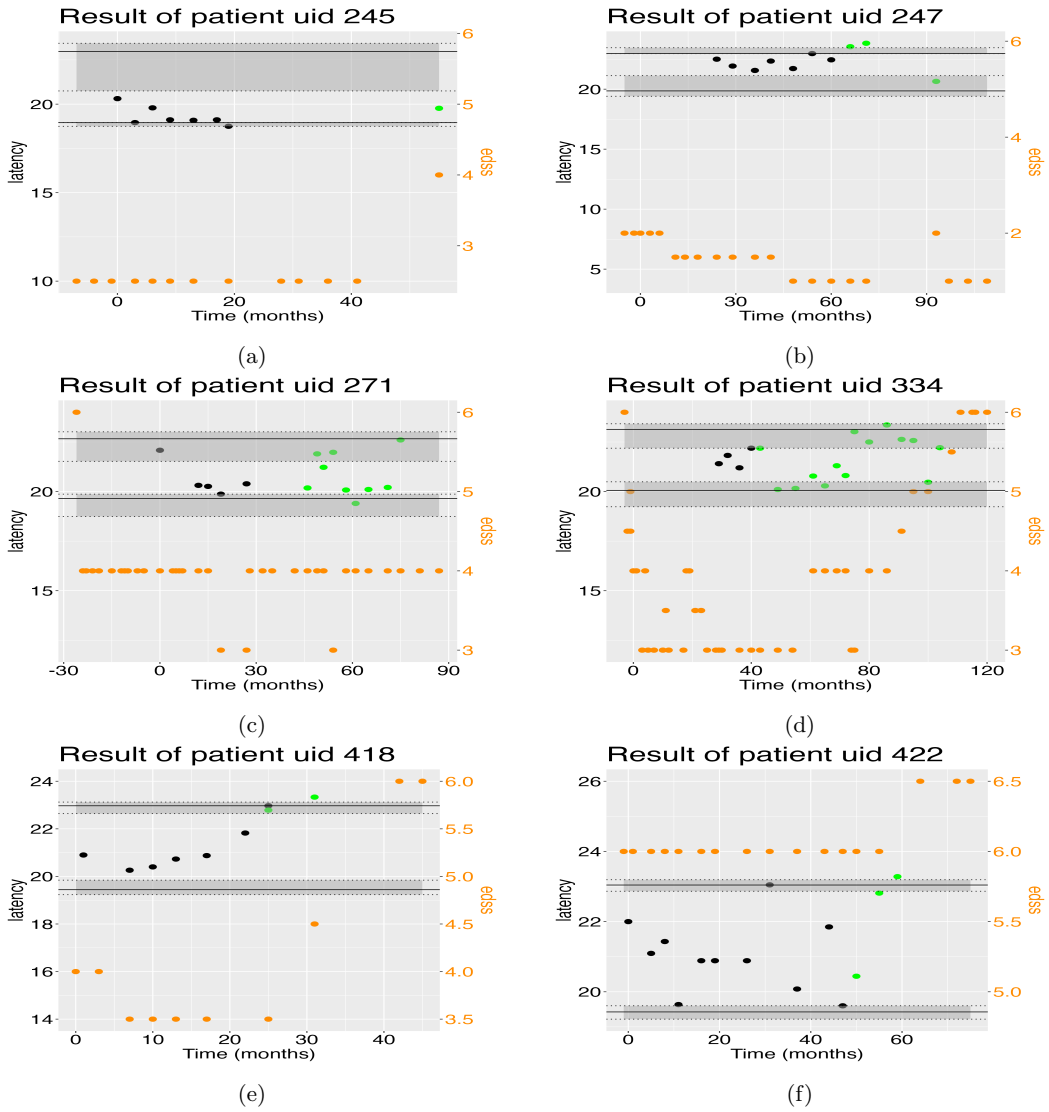


Figure 82: A subset of PwMS from the right hand data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points present the EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression cannot be detected earlier by IRI than EDSS score due to no data available during the time intervla of 2 years
(b) Progression can be detected at 98 months based on EDSS score, and at 70 months based on IRI.
(c) Progression cannot be detected earlier by IRI than EDSS score due to no data available during the time intervla of 2 years.
(d) Progression can be detected at 72 months based on EDSS score, and at 46 months based on IRI.
(e) Progression can be detected at 45 months based on EDSS score, and at 25 months based on IRI.
(f) Progression can be detected at 76 months based on EDSS score, and at 56 months based on IRI.

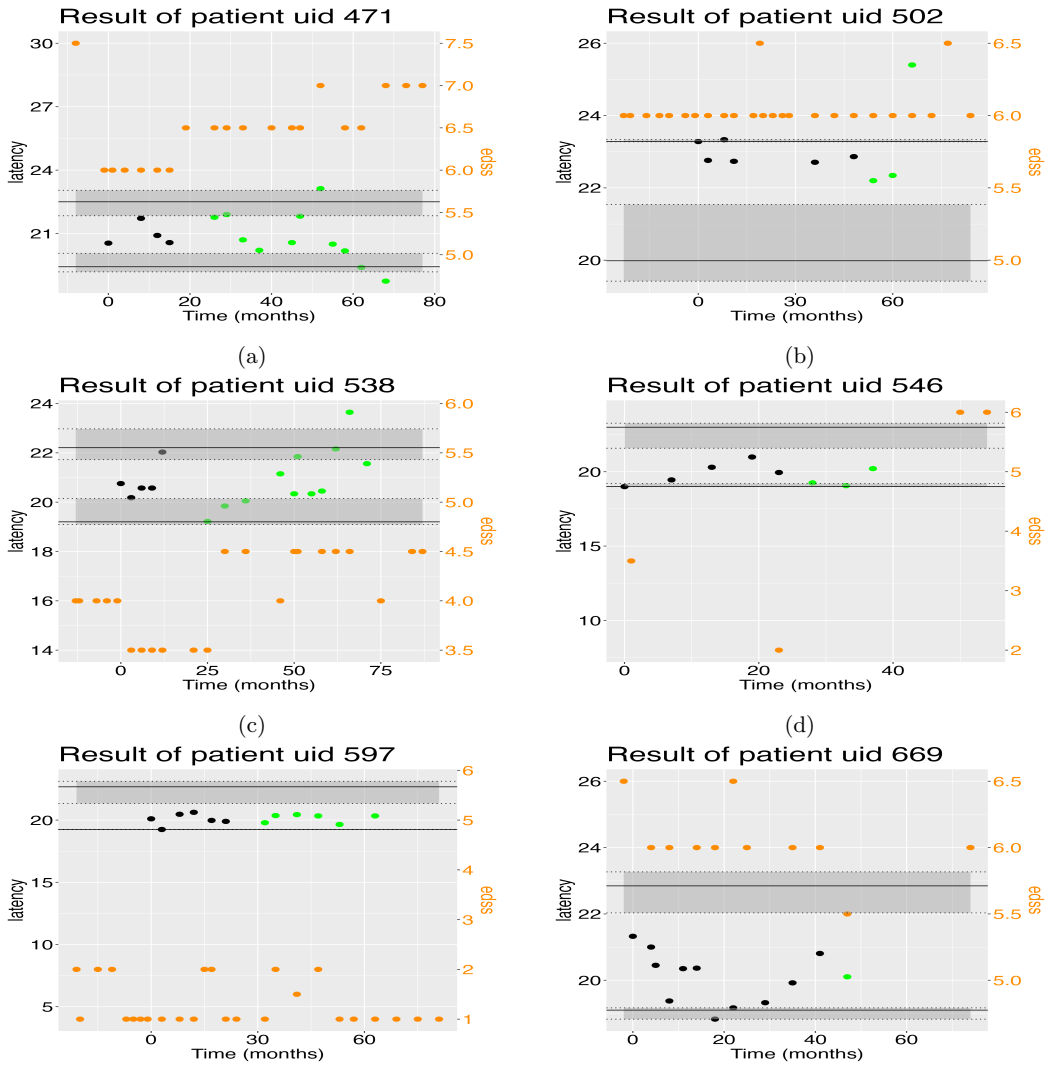


Figure 83: A subset of PwMS from the right hand data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression can be detected at 48 months based on EDSS score, and at 33 months based on IRI.
(b) Progression can be detected at 99 months based on EDSS score, and at 89 months based on IRI.
(c) Progression can be detected at 49 months based on EDSS score, and at 38 months based on IRI.
(d) Progression can be detected at 50 months based on EDSS score, and at 28 months based on IRI.
(e) Progression cannot detected earlier by IRI than by EDSS score
(f) Progression cannot be detected earlier by IRI than EDSS score due to no data available during the time intervla of 2 years.

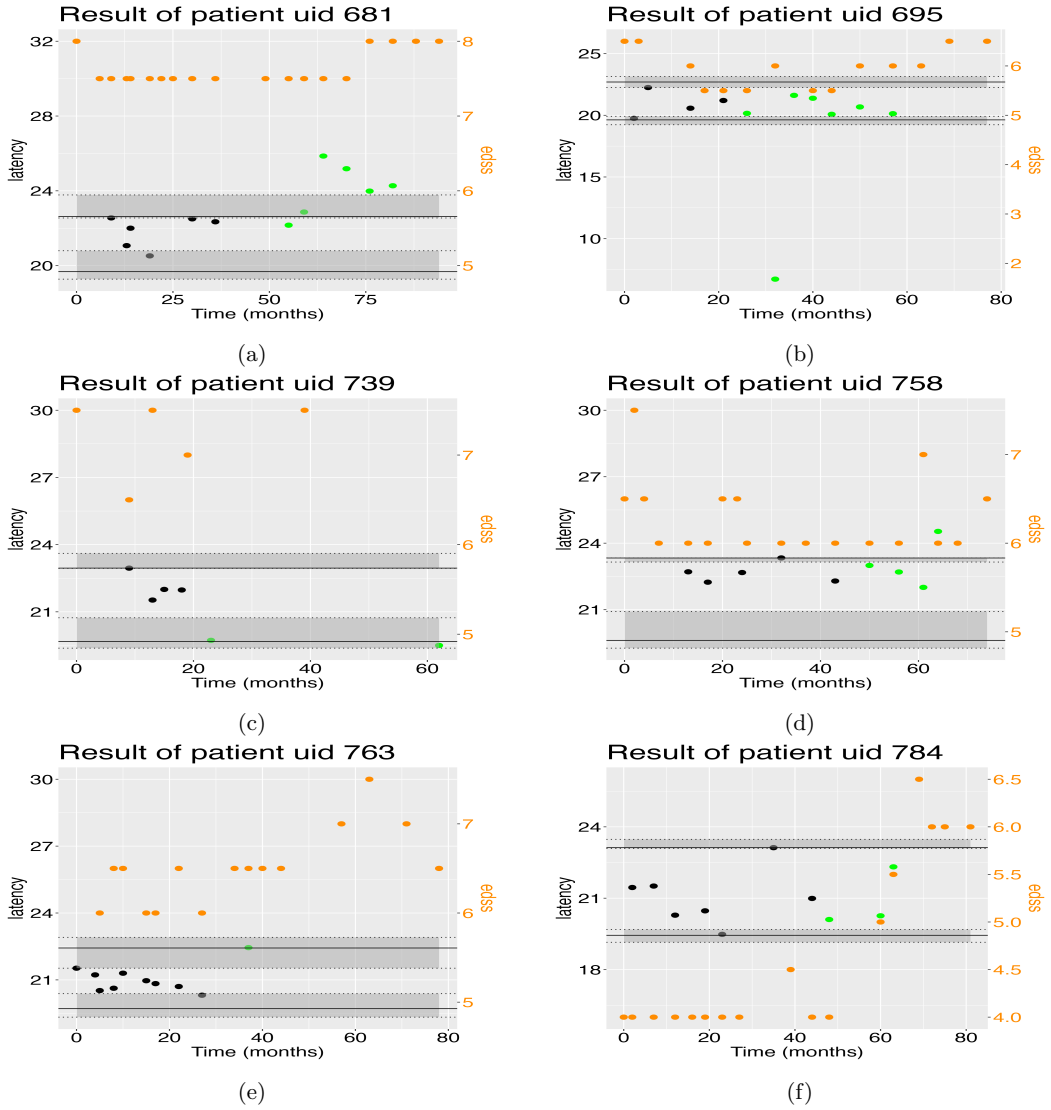


Figure 84: A subset of PwMS from the right hand data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression can be detected at 76 months based on EDSS score, and at 59 months based on IRI.
(b) Progression can be detected at 50 months based on EDSS score, and at 32 months based on IRI.
(c) Progression can be detected at 39 months based on EDSS score, and at 23 months based on IRI.
(d) Progression can be detected at 74 months based on EDSS score, and at 64 months based on IRI.
(e) Progression can be detected at 57 months based on EDSS score, and at 37 months based on IRI.
(f) Progression cannot be detected earlier by IRI than by EDSS score.

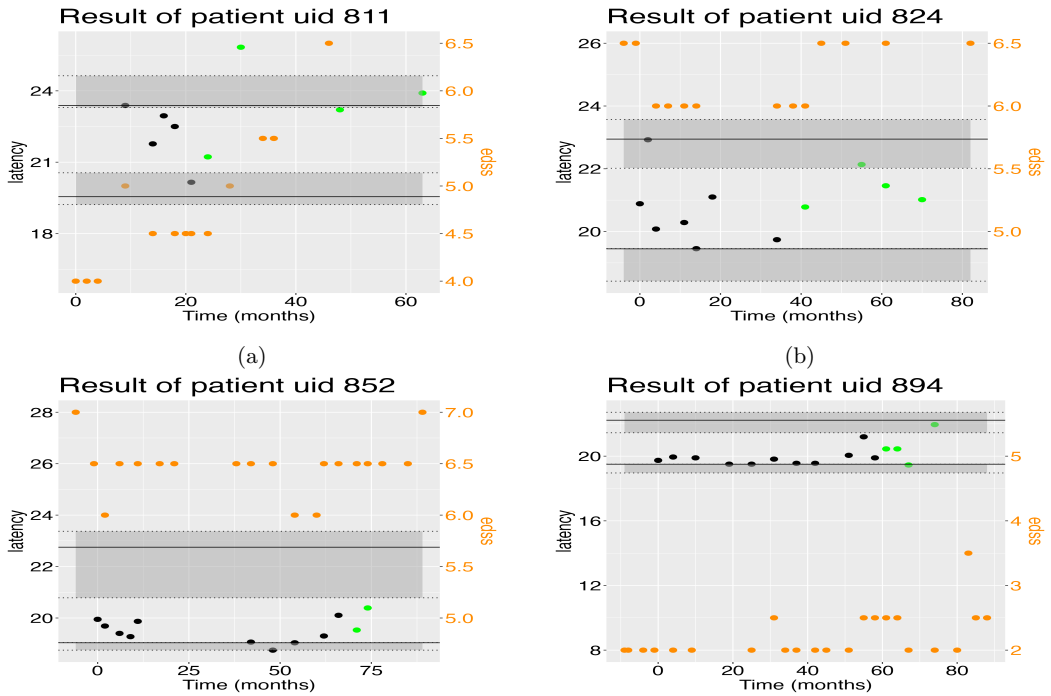


Figure 85: A subset of PwMS from the right hand data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression can be detected at 46 months based on EDSS score, and at 30 months based on IRI.
(b) Progression can be detected at 64 months based on EDSS score, and at 59 months based on IRI.
(c) Progression cannot be detected earlier by IRI than by EDSS score
(d) Progression can be detected at 92 months based on EDSS score, and at 75 months based on IRI.

M Confidence Interval of IRI and full EDSS for right foot data

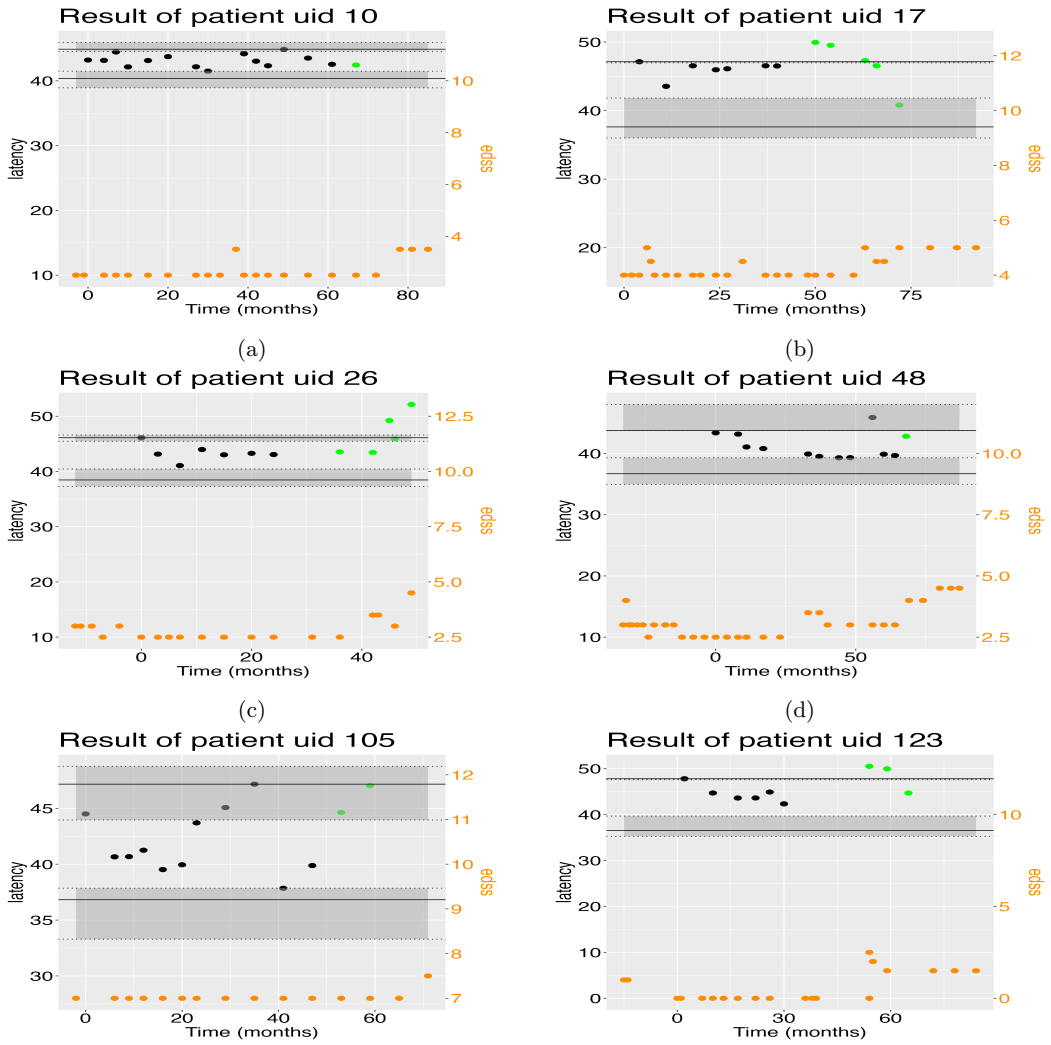


Figure 86: A subset of PwMS from the right foot data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression cannot be detected earlier by IRI than by EDSS score due to no data during the time interval of 2 years.
(b) Progression can be detected at 72 months based on EDSS score, and at 50 months based on IRI.
(c) Progression can be detected at 60 months based on EDSS score, and at 56 months based on IRI.
(d) Progression cannot be detected earlier by IRI than by EDSS score
(e) Progression can be detected at 73 months based on EDSS score, and at 55 months based on IRI.
(f) Progression can be detected at 73 months based on EDSS score, and at 68 months based on IRI.

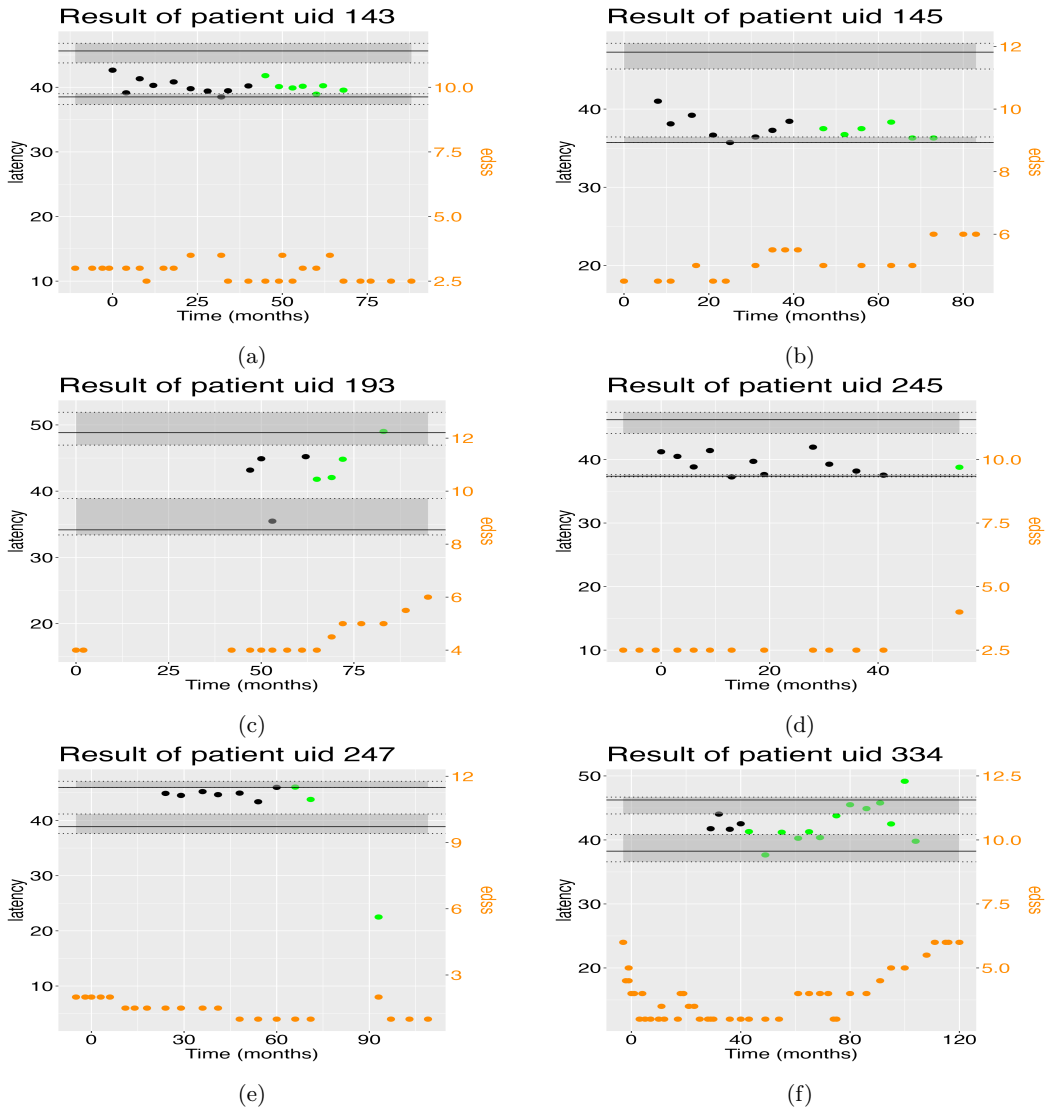


Figure 87: A subset of PwMS from the right foot data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression can be detected at 74 months based on EDSS score, and at 70 months based on IRI.
(b) Progression can be detected at 73 months based on EDSS score, and at 68 months based on IRI.
(c) Progression can be detected at 89 months based on EDSS score, and at 83 months based on IRI.
(d) Progression cannot be detected earlier by IRI than by EDSS score due to no data available during the time interval of 2 years.
(e) Progression can be detected at 98 months based on EDSS score, and at 70 months based on IRI.
(f) Progression can be detected at 72 months based on EDSS score, and at 52 months based on IRI.

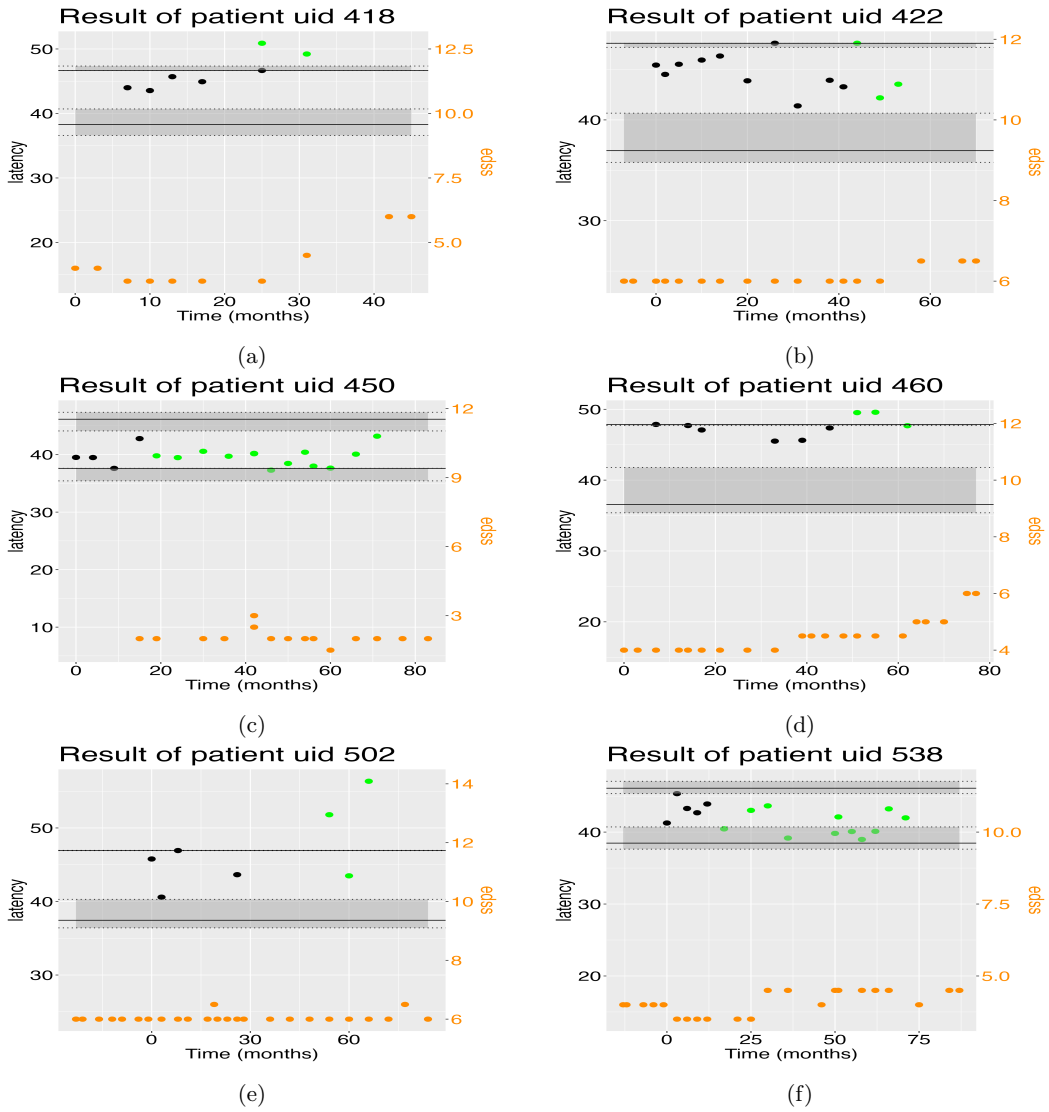


Figure 88: A subset of PwMS from the right foot data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression can be detected at 45 months based on EDSS score, and at 25 months based on IRI.
(b) Progression can be detected at 76 months based on EDSS score, and at 51 months based on IRI.
(c) Progression cannot be detected earlier by IRI than by EDSS score.
(d) Progression can be detected at 75 months based on EDSS score, and at 51 months based on IRI.
(e) Progression can be detected at 99 months based on EDSS score, and at 77 months based on IRI.
(f) Progression can be detected at 49 months based on EDSS score, and at 29 months based on IRI.

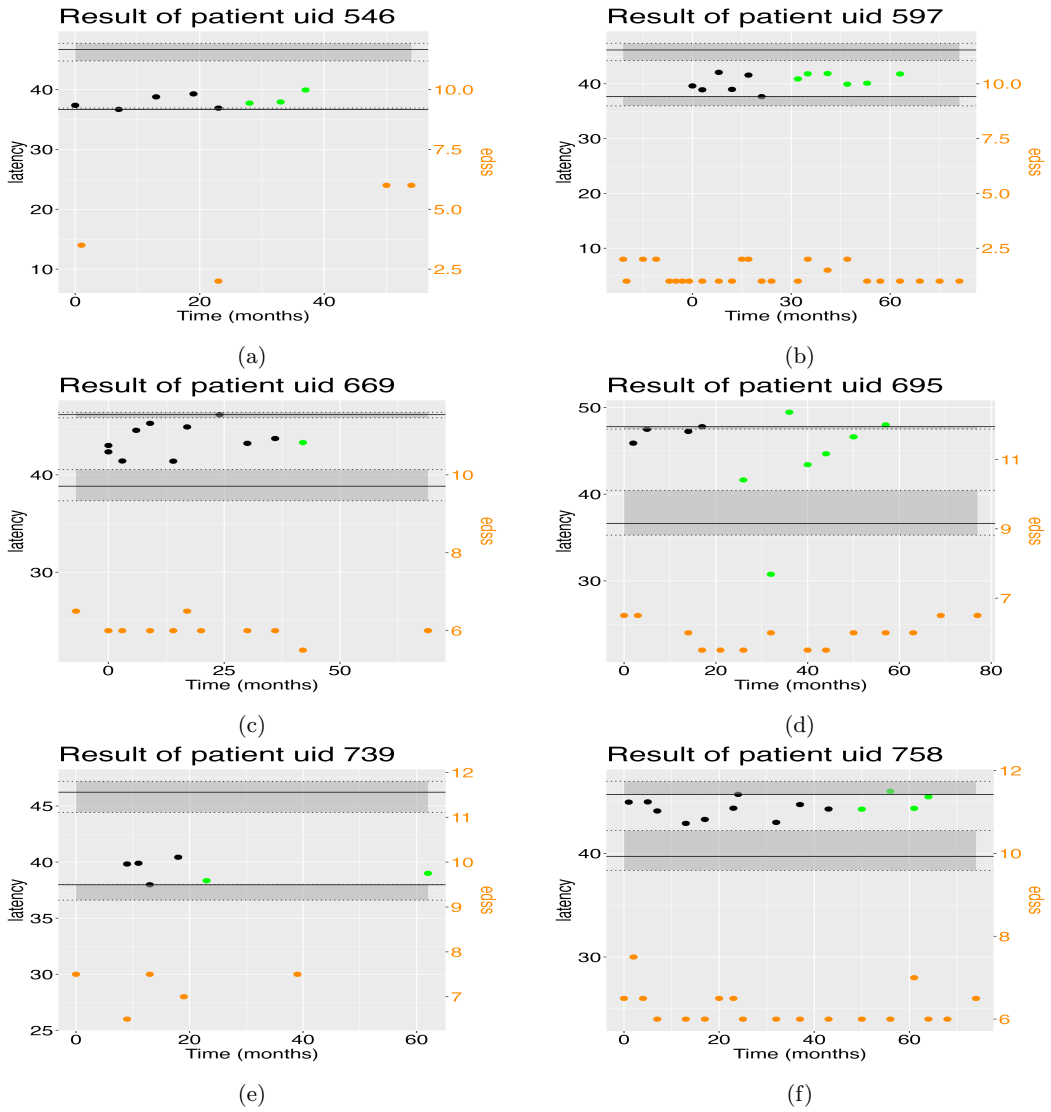


Figure 89: A subset of PwMS from the right foot data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression cannot be detected earlier by IRI than by EDSS score.
(b) Progression cannot be detected earlier by IRI than by EDSS score.
(c) Progression cannot be detected earlier by IRI than by EDSS score due to no data available during the time interval of 2 years.
(d) Progression can be detected at 50 months based on EDSS score, and at 32 months based on IRI.
(e) Progression cannot be detected earlier by IRI than by EDSS score.
(f) Progression can be detected at 74 months based on EDSS score, and at 56 months based on IRI.

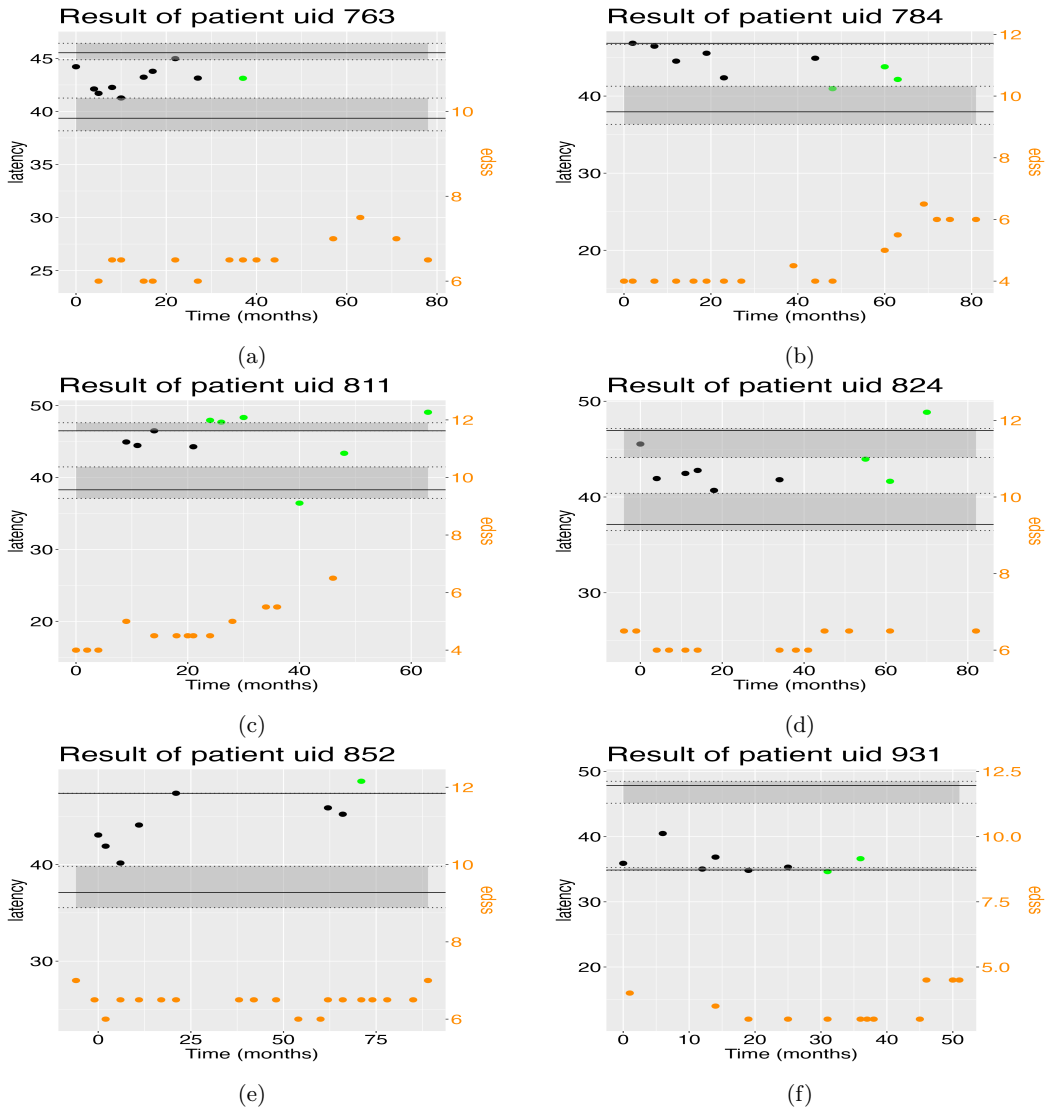
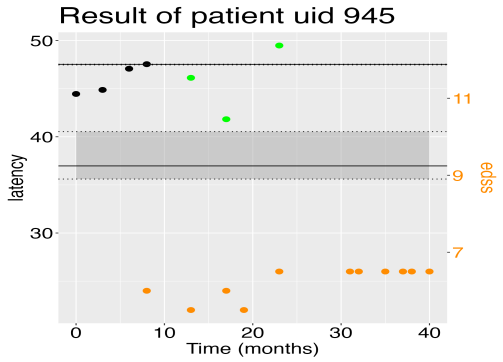


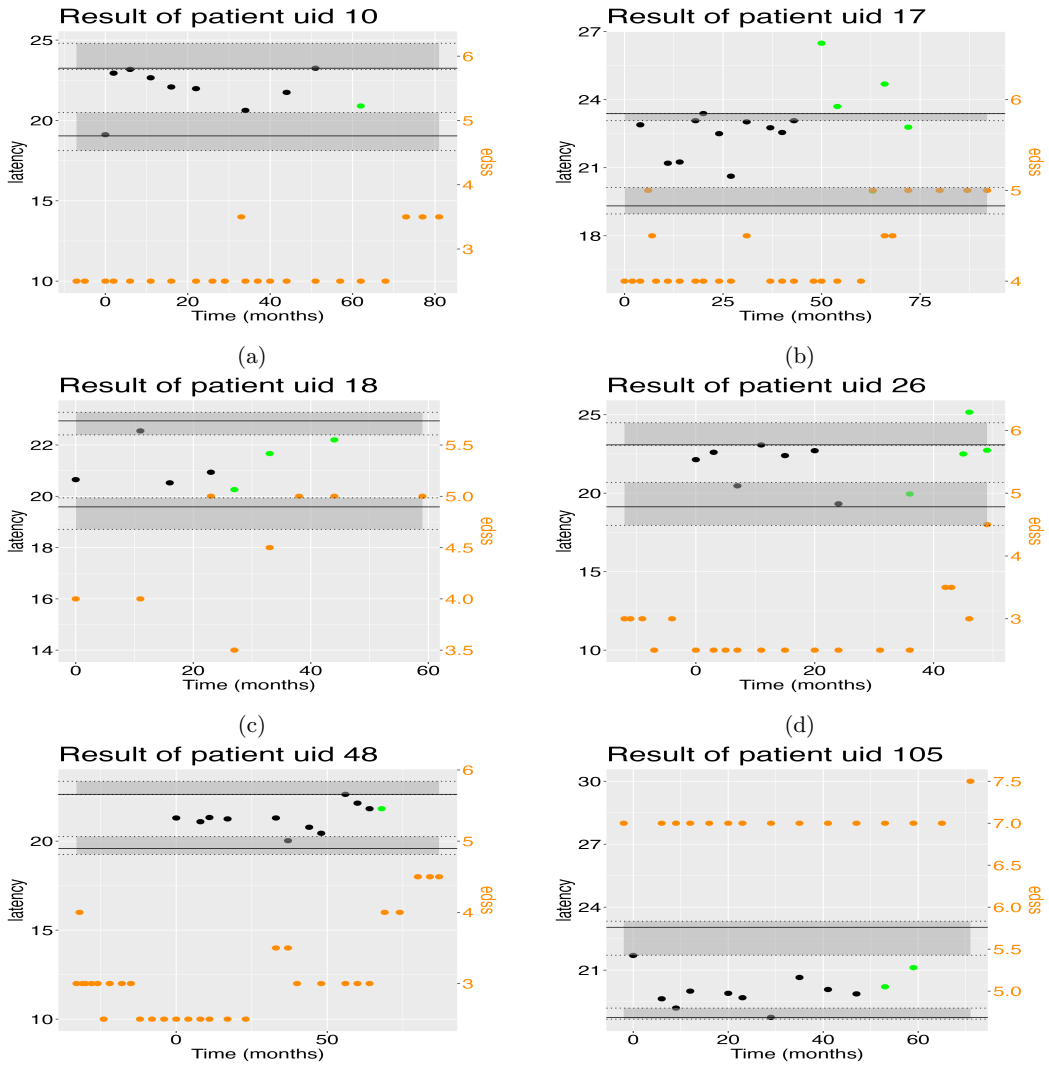
Figure 90: A subset of PwMS from the right foot data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression cannot be detected earlier by IRI than by EDSS score.
(b) Progression can be detected at 72 months based on EDSS score, and at 48 months based on IRI.
(c) Progression can be detected at 46 months based on EDSS score, and at 24 months based on IRI.
(d) Progression can be detected at 64 months based on EDSS score, and at 59 months based on IRI.
(e) Progression can be detected at 95 months based on EDSS score, and at 77 months based on IRI.
(f) Progression can be detected at 50 months based on EDSS score, and at 31 months based on IRI.



(a)

Figure 91: A subset of PwMS from the right foot data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points present the EDSS score. The progression of MS can be detected earlier using IRI than EDSS score. (a) Progression can be detected at 32 months based on EDSS score, and at 23 months based on IRI.

N Confidence Interval of IRI and full EDSS for left hand data



(a) (b) (c) (d) (e) (f)

Figure 92: A subset of PwMS from the left hand data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression cannot be detected earlier by IRI than by EDSS score.
(b) Progression can be detected at 72 months based on EDSS score, and at 50 months based on IRI.
(c) Progression cannot be detected earlier by IRI than by EDSS score.
(d) Progression can be detected at 60 months based on EDSS score, and at 47 months based on IRI.
(e) Progression cannot be detected earlier by IRI than by EDSS score.
(f) Progression cannot be detected earlier by IRI than by EDSS score.

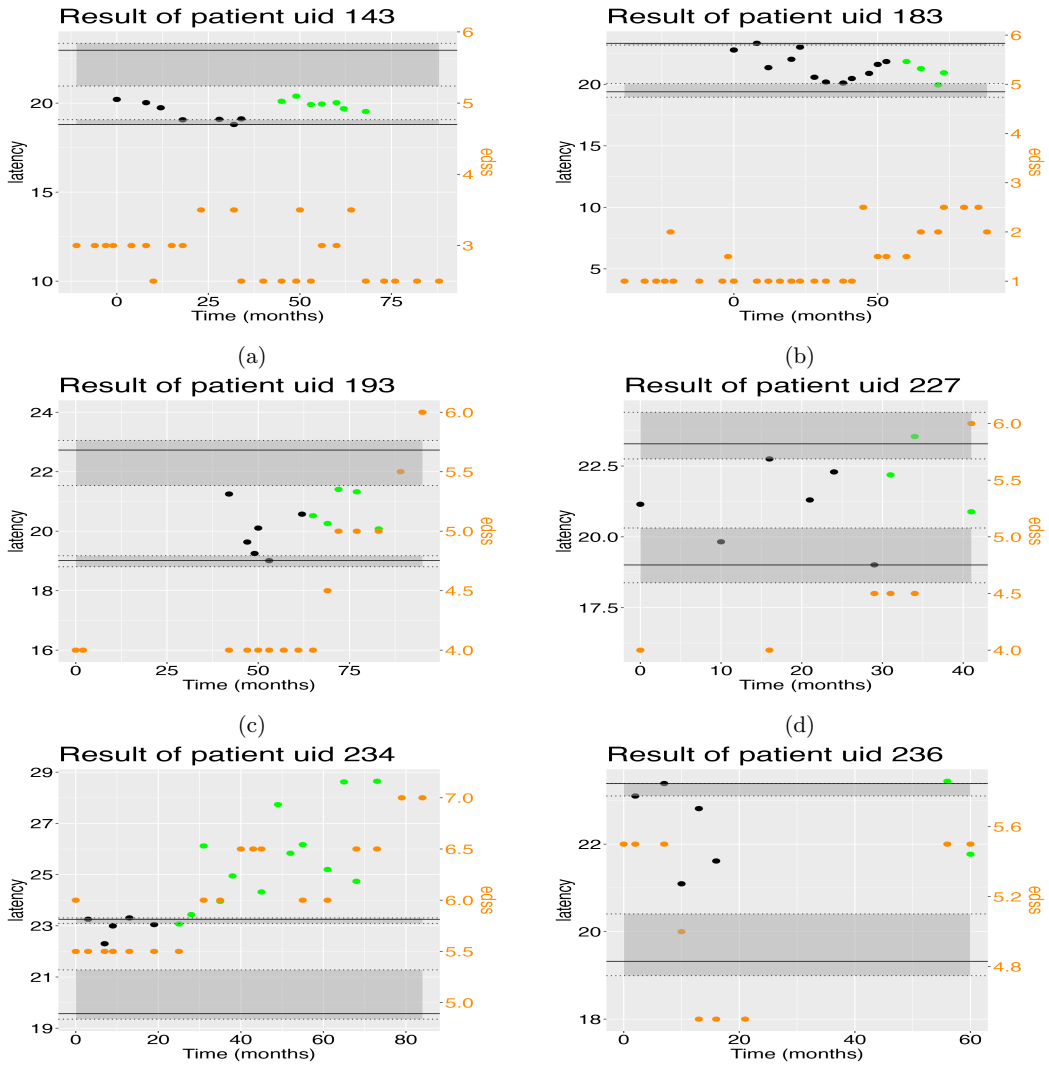


Figure 93: A subset of PwMS from the left hand data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points present the EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression cannot be detected earlier by IRI than by EDSS score.
(b) Progression can be detected at 122 months based on EDSS score, and at 108 months based on IRI.
(c) Progression cannot be detected earlier by IRI than by EDSS score.
(d) Progression can be detected at 41 months based on EDSS score, and at 34 months based on IRI.
(e) Progression can be detected at 45 months based on EDSS score, and at 25 months based on IRI.
(f) Progression cannot be detected earlier by IRI than by EDSS score due to no data available during the time interval of 2 years.

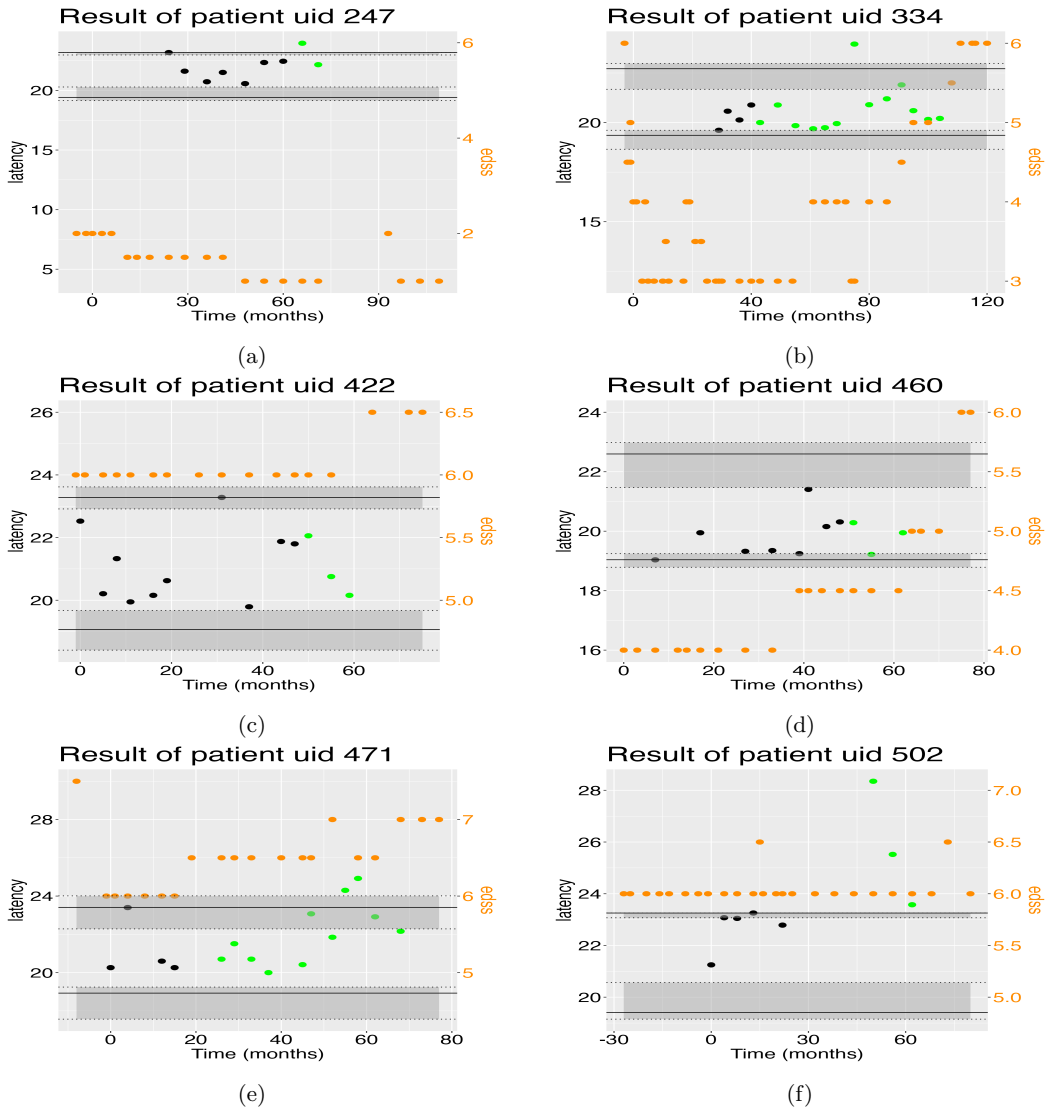


Figure 94: A subset of PwMS from the left hand data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression can be detected at 98 months based on EDSS score, and at 70 months based on IRI.
(b) Progression cannot be detected earlier by IRI than by EDSS score.
(c) Progression cannot be detected earlier by IRI than by EDSS score.
(d) Progression can be detected at 75 months based on EDSS score, and at 55 months based on IRI.
(e) Progression cannot be detected earlier by IRI than by EDSS score.
(f) Progression can be detected at 99 months based on EDSS score, and at 77 months based on IRI.

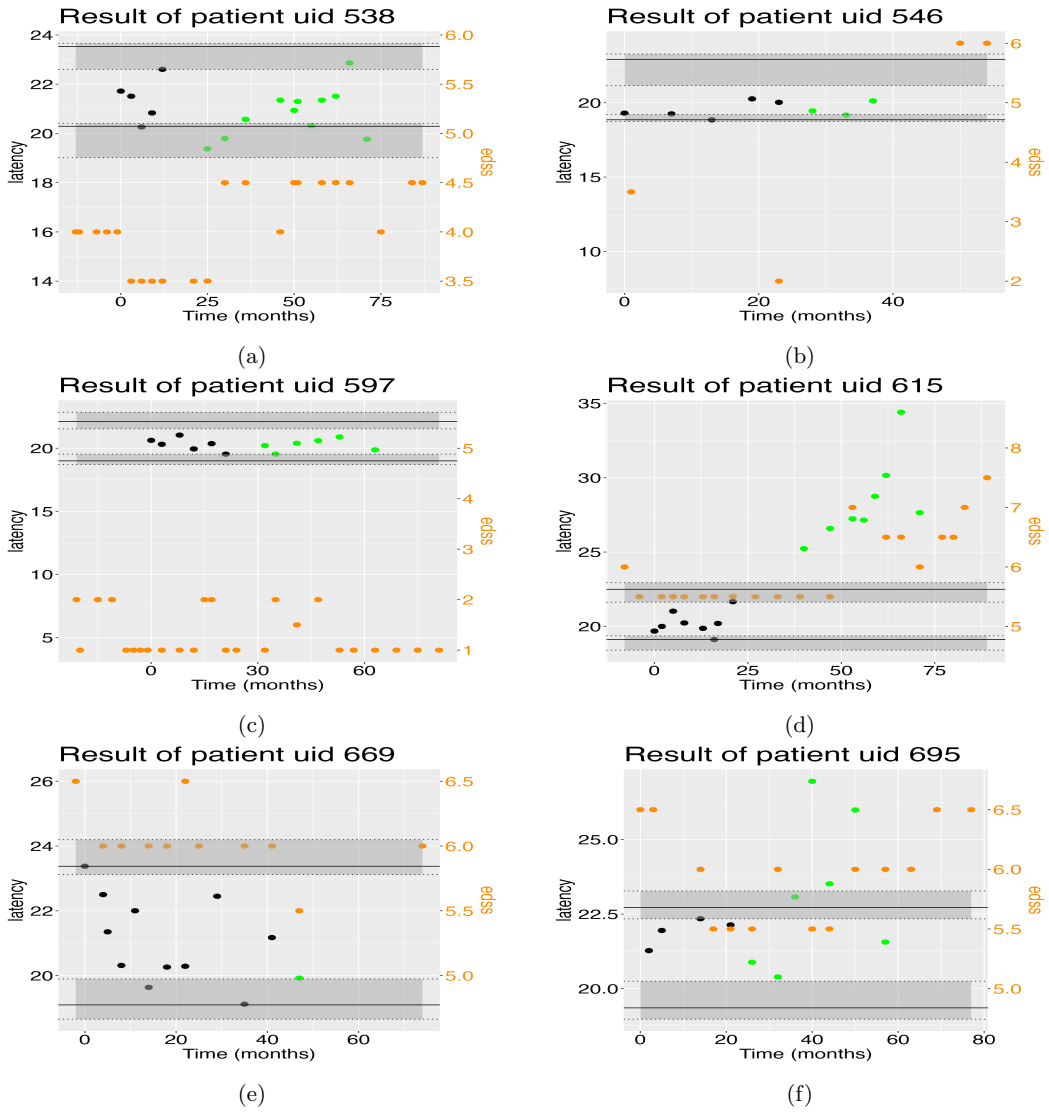


Figure 95: A subset of PwMS from the left hand data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points present the EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression can be detected at 49 months based on EDSS score, and at 38 months based on IRI.
(b) Progression can be detected at 50 months based on EDSS score, and at 33 months based on IRI.
(c) Progression can be detected at 68 months based on EDSS score, and at 56 months based on IRI.
(d) Progression can be detected at 73 months based on EDSS score, and at 47 months based on IRI.
(e) Progression cannot be detected earlier by IRI than by EDSS score.
(f) Progression can be detected at 50 months based on EDSS score, and at 36 months based on IRI.

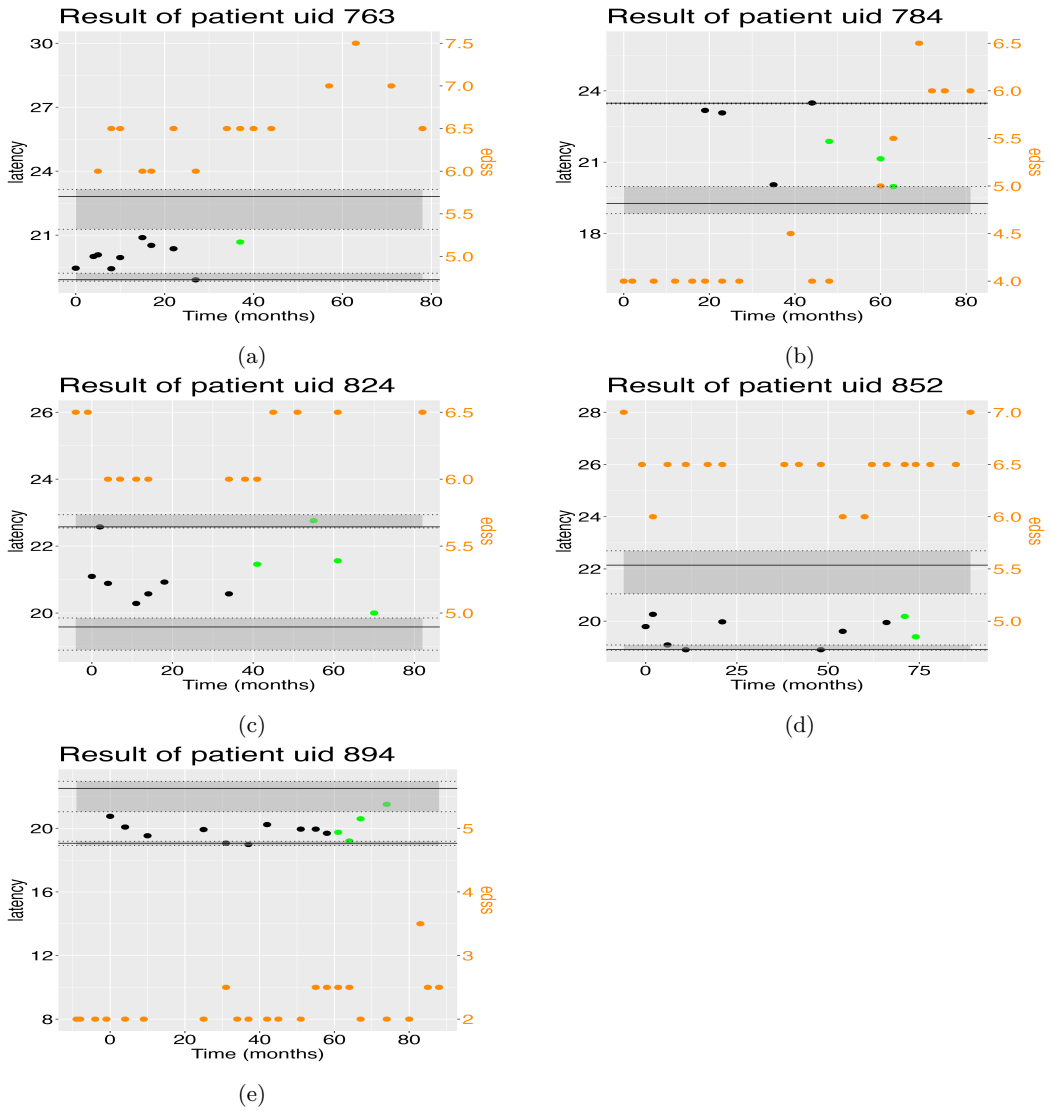
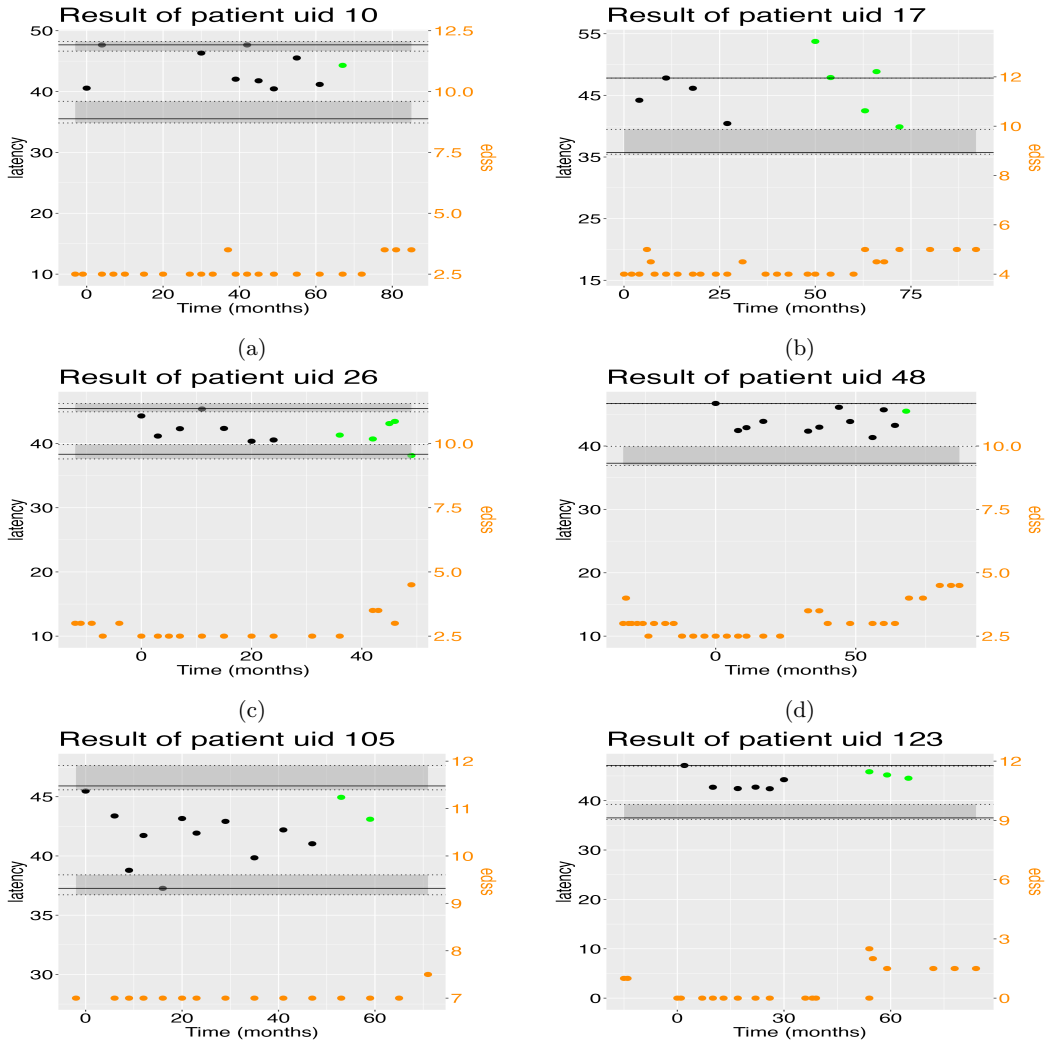


Figure 96: A subset of PwMS from the left hand data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression cannot be detected earlier by IRI than by EDSS score.
(b) Progression can be detected at 72 months based on EDSS score, and at 63 months based on IRI.
(c) Progression can be detected at 64 months based on EDSS score, and at 59 months based on IRI.
(d) Progression cannot be detected earlier by IRI than by EDSS score.
(e) Progression can be detected at 76 months based on EDSS score, and at 59 months based on IRI.
(f) Progression can be detected at 92 months based on EDSS score, and at 73 months based on IRI.

O Confidence Interval of IRI and full EDSS left foot data



(a) (b) (c) (d) (e) (f)

Figure 97: A subset of PwMS from the left foot data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points present the EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression cannot be detected earlier by IRI than by EDSS score.
(b) Progression can be detected at 72 months based on EDSS score, and at 50 months based on IRI.
(c) Progression cannot be detected earlier by IRI than by EDSS score.
(d) Progression cannot be detected earlier by IRI than by EDSS score.
(e) Progression cannot be detected earlier by IRI than by EDSS score.
(f) Progression cannot be detected earlier by IRI than by EDSS score.

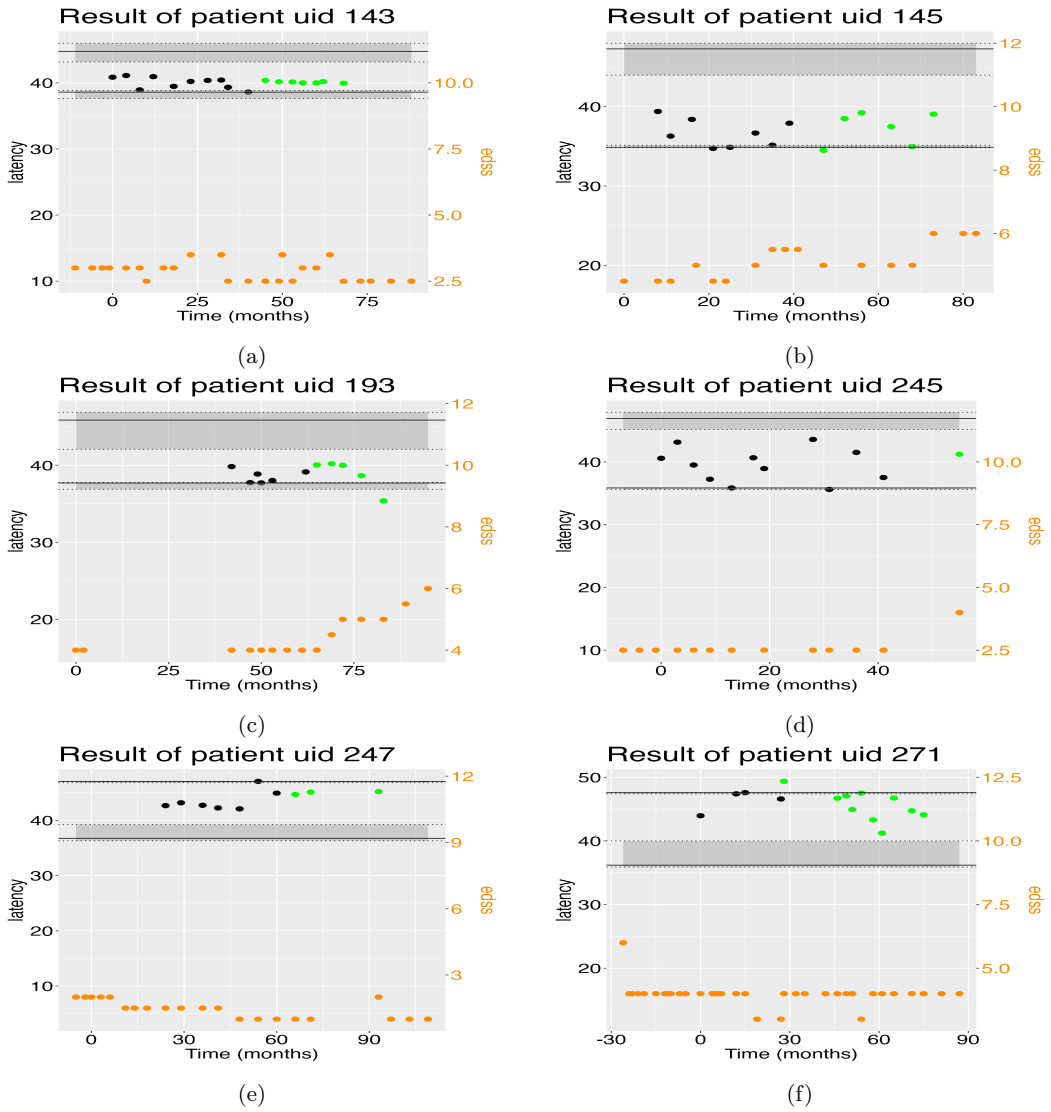


Figure 98: A subset of PwMS from the left foot data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression cannot be detected earlier by IRI than by EDSS score.
(b) Progression can be detected at 73 months based on EDSS score, and at 47 months based on IRI.
(c) Progression can be detected at 89 months based on EDSS score, and at 83 months based on IRI.
(d)Progression cannot be detected earlier by IRI than by EDSS score.
(e) Progression cannot be detected earlier by IRI than by EDSS score.
(f) Progression can be detected at 76 months based on EDSS score, and at 53 months based on IRI.

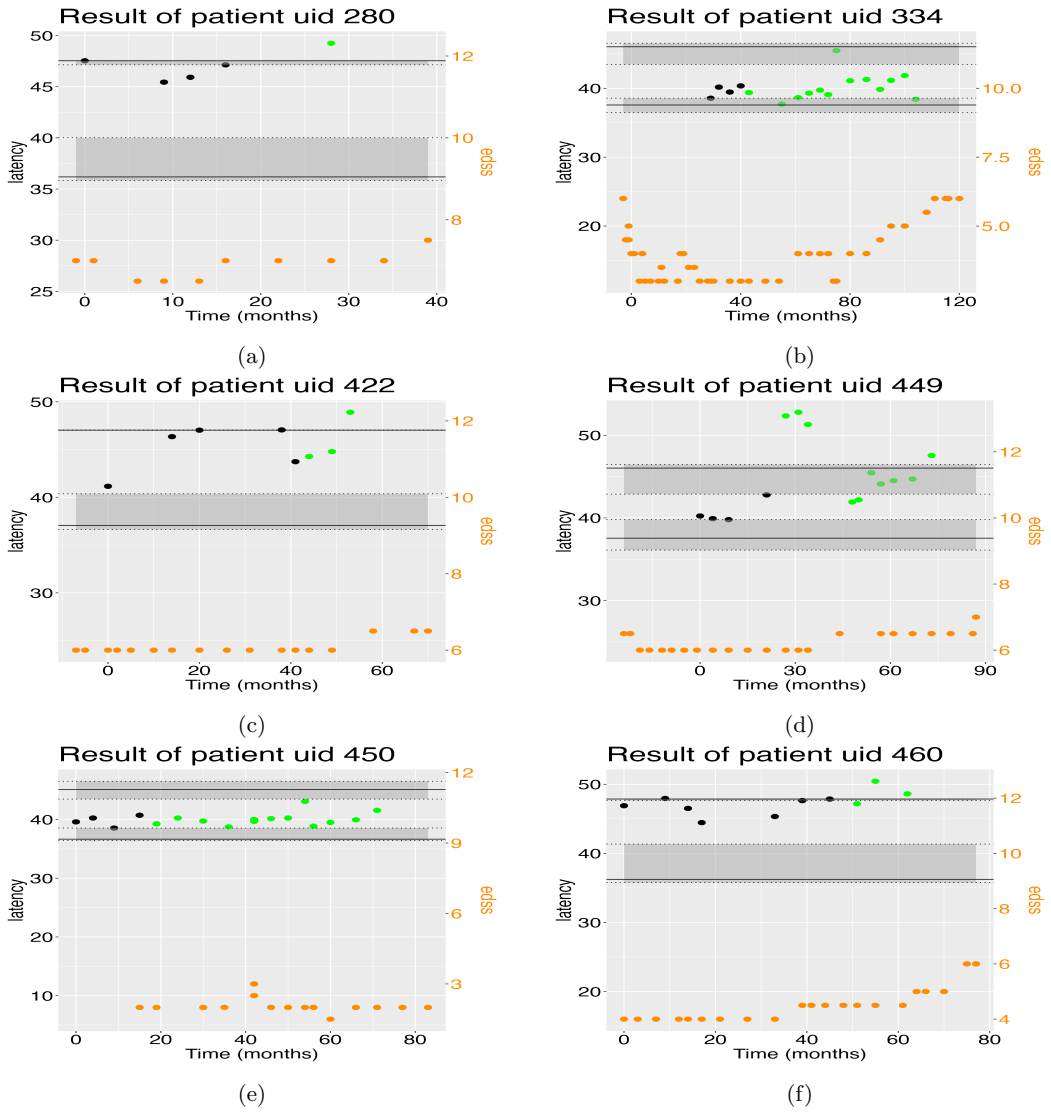


Figure 99: A subset of PwMS from the left foot data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression can be detected at 40 months based on EDSS score, and at 29 months based on IRI.
(b) Progression can be detected at 72 months based on EDSS score, and at 57 months based on IRI.
(c) Progression can be detected at 76 months based on EDSS score, and at 60 months based on IRI.
(d) Progression can be detected at 67 months based on EDSS score, and at 251 months based on IRI.
(e)Progression cannot be detected earlier by IRI than by EDSS score.
(f) Progression can be detected at 75 months based on EDSS score, and at 55 months based on IRI.

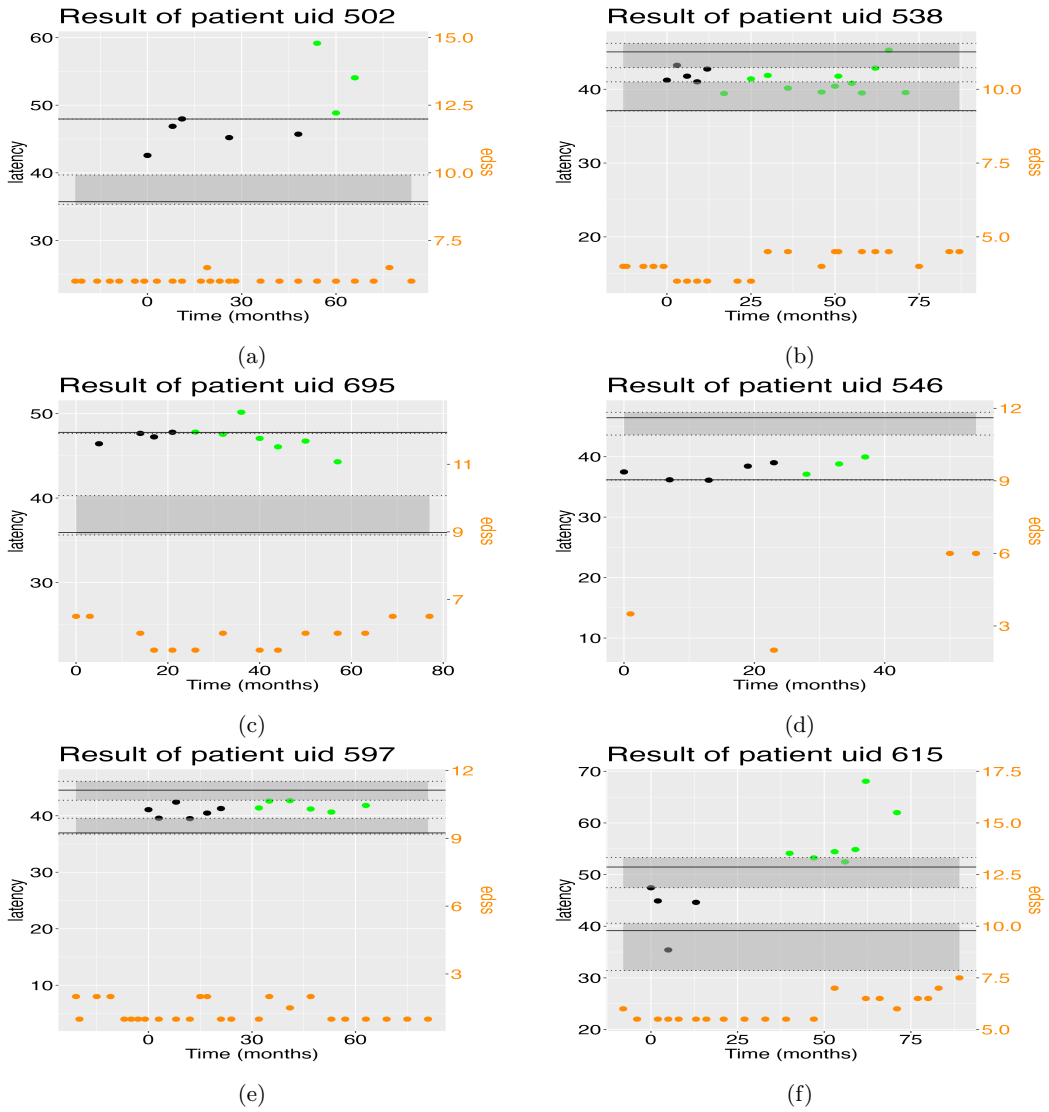


Figure 100: A subset of PwMS from the left foot data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points present the EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression can be detected at 99 months based on EDSS score, and at 77 months based on IRI.
(b) Progression can be detected at 49 months based on EDSS score, and at 29 months based on IRI.
(c) Progression can be detected at 50 months based on EDSS score, and at 26months based on IRI.
(d) Progression cannot be detected earlier by IRI than by EDSS score.
(e) Progression can be detected at 68 months based on EDSS score, and at 56 months based on IRI.
(f) Progression can be detected at 73 months based on EDSS score, and at 47 months based on IRI.

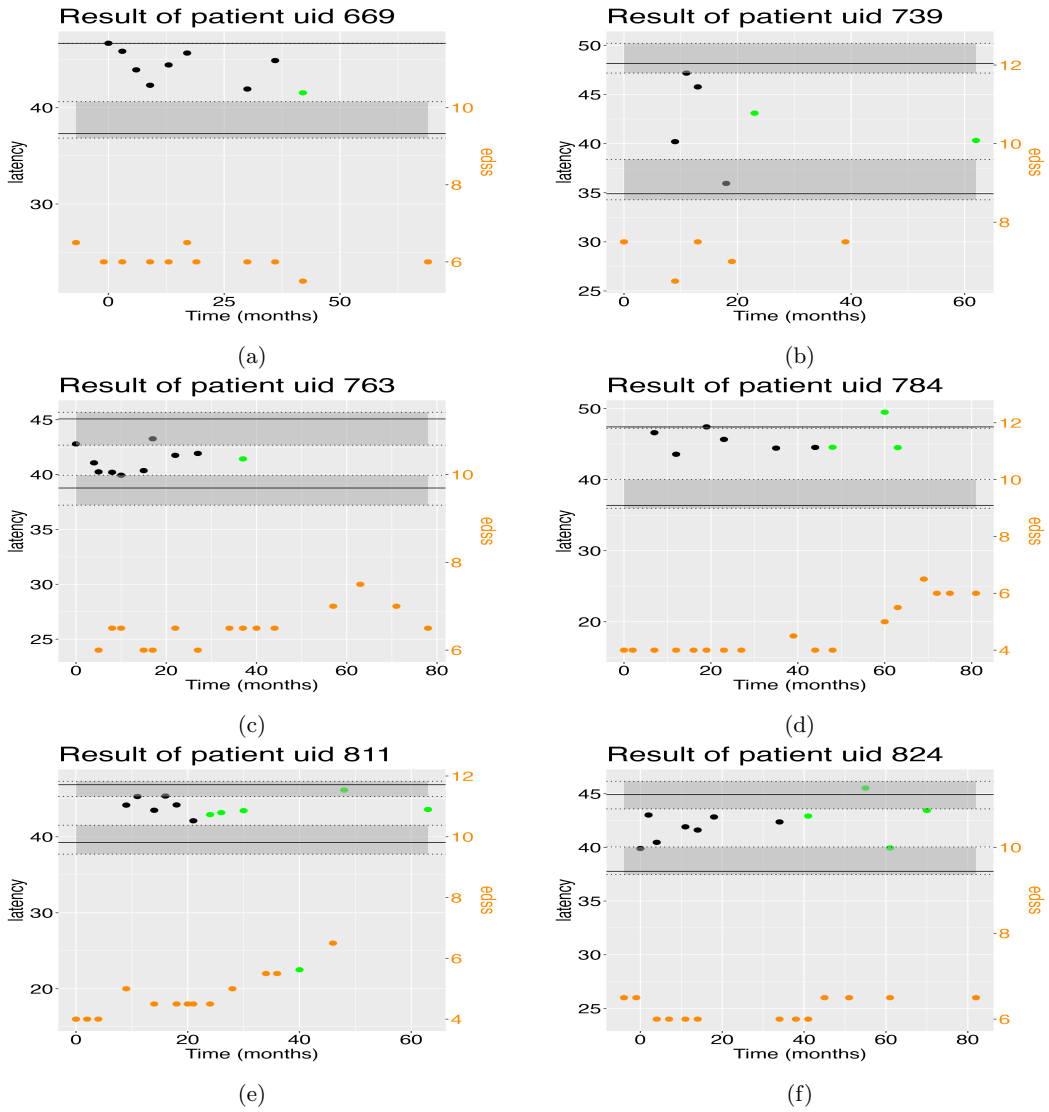


Figure 101: A subset of PwMS from the left foot data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points presents EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression cannot be detected earlier by IRI than by EDSS score.
(b)Progression cannot be detected earlier by IRI than by EDSS score.
(c) Progression cannot be detected earlier by IRI than by EDSS score.
(d) Progression can be detected at 72 months based on EDSS score, and at 60 months based on IRI.
(e) Progression can be detected at 46 months based on EDSS score, and at 40 months based on IRI.
(f) Progression can be detected at 64 months based on EDSS score, and at 59 months based on IRI.

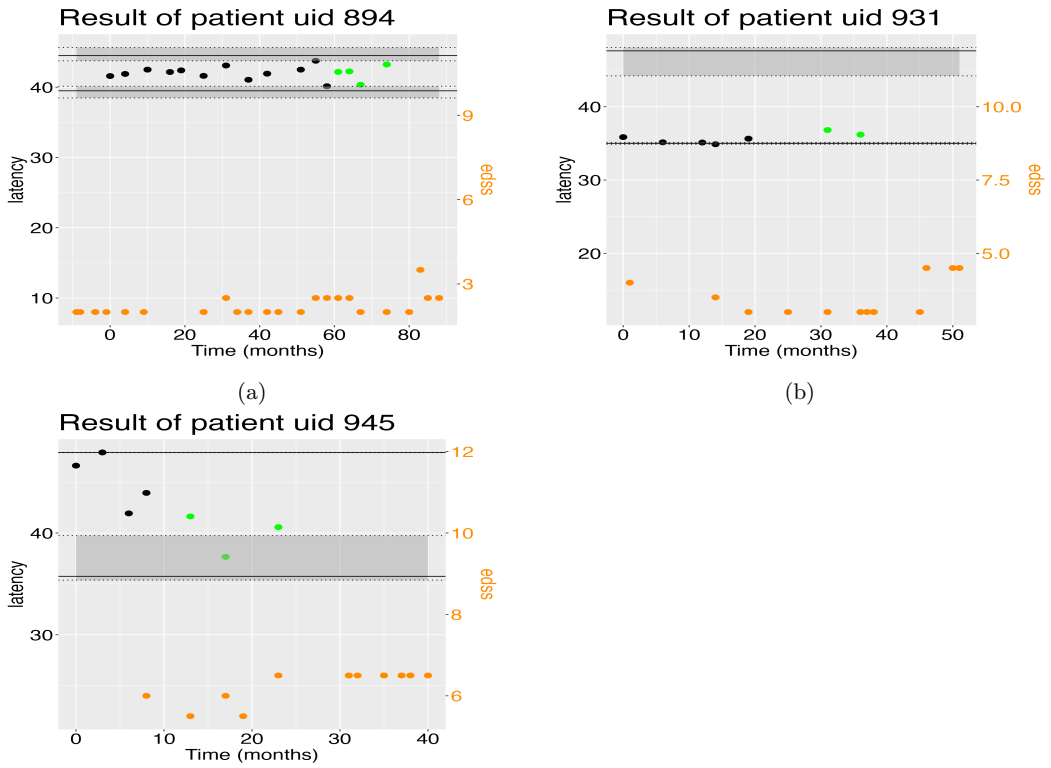


Figure 102: A subset of PwMS from the left foot data. The black solid line presents the lower and the upper boundaries of IRI, the black points present the stable data, the green points present the future data, the gray zone presents the 95% CI of the lower (upper) bound. The orange points present the EDSS score. The progression of MS can be detected earlier using IRI than EDSS score.
(a) Progression cannot be detected earlier by IRI than by EDSS score.
(b) Progression cannot be detected earlier by IRI than by EDSS score.
(c) Progression can be detected at 32 months based on EDSS score, and at 17 months based on IRI.

P Selected Code

```
setwd("/Users/macbookair/Documents/UHasselt:Master-Thesis/Coding
/R-Code/Data_Analysis_April/")
data_foot_right <- read.csv("training_data_foot_right.csv")
test_data_foot_right <- read.csv("test_data_foot_right.csv")
number_obserb <- nrow(data_foot_right) # 2419
list_id_outlier <- c()
data_foot_right_new <- data_foot_right[FALSE, ]
# create an df that contains same col as in data_foot_right
for (i in 1:number_obserb)
{
  if ((data_foot_right$marker1[i] >= 150) |(data_foot_right$marker1[i] <= 0))
  {
    list_id_outlier <- c(list_id_outlier, data_foot_right$patient_uid[i])
  }
  else
  {
    data_foot_right_new <- rbind(data_foot_right_new, data_foot_right[i,])
  }
}

data_foot_right_median_new <- data_foot_right_new %>%
dplyr::group_by(patient_uid, visit_date) %>%
dplyr::summarise(mean_marker1=mean(marker1)) %>%
ungroup()
res_foot_no_stab <-jqm(db=data_foot_right_median_new_3,
          alpha=0.05)
          res <- res_foot_no_stab
patient_uid <- unique(data_foot_right_median_new_3$subject)

foot_right.jqm<-cbind.data.frame(res$beta0+res$u+res$z*res$beta1,
                                 res$beta0+res$u+res$z*res$beta2,
                                 seq(1:length(unique(data_foot_right_median_new_3$
                                 subject))), patient_uid)
colnames(foot_right.jqm)<-c("low","up","id", "patient_uid")
data_foot_right_median_new_3 <-transform(data_foot_right_median_new_3,
ID = as.numeric(factor(subject)))
id_list <- unique(data_foot_right_median_new_3$subject)
test_data_foot_right_new <- test_data_foot_right %>%
```

```
filter(patient_uid %in% id_list)

test_data_foot_right_median_new <- test_data_foot_right_new %>%
dplyr::group_by(patient_uid, visit_date) %>%
dplyr::summarise(mean_marker1=median(marker1)) %>%
ungroup()
```