a linear system approach. Such approach has not been successful in explaining and/or predicting RRIs satisfactorily.

**Objectives** To develop and validate an artificial intelligence (AI) algorithm in order to identify RRI risk profiles in recreational runners.

**Design** Mathematical model.

**Settings** São Paulo, Brazil.

**Participants** 191 recreational runners.

**Assessment of Risk Factors** This was a 3-step AI study using data from a prospective cohort study. In step 1, variable selection and exploratory analyses were conducted in the original (n=191) and simulated data (n=5000). In step 2, the AI algorithm was developed using self-organising maps, k-means and probabilistic neural networks. The algorithm was trained in 80% (n=4000) of the simulated data, and validated using the remaining 20% (n=1000). Characterisation of RRI risk profiles was performed in step 3.

**Main Outcome Measures** RRI risk profiles were established based on the groups created by the developed algorithm. Descriptive analyses were performed to summarise the risk profiles.

**Results** The variables with greatest influence in the algorithm were: sex; running intensity; history of RRIs; and current musculoskeletal discomfort related to running. Five groups were suggested by the algorithm. Male runners reporting previous RRIs and running in low-to-moderate intensities (≥6 min/km) were at the highest risk of RRIs. Male runners reporting previous RRIs and running in high intensities (3 to 5 min/km) in about 32.1% of the time were at the lowest risk of RRIs. The classification accuracy of the algorithm presented a median of 99.6% (interquartile range: 99.5% to 99.8%).

**Conclusions** A non-linear system approach using AI and machine learning techniques were successful in developing an RRI risk profile algorithm for recreational runners.