Antiretroviral therapy (ART) effectively controls HIV infection, suppressing HIV viral loads. Typically suspension of therapy is rapidly followed by rebound of viral loads to high, pre-therapy levels. However, case reports suggest that initiating ART early after infection may delay viral rebound, for months, years, or maybe permanently, after ART suspension. We will discuss our multi-type, branching process model to gain insight into these post-treatment dynamics. Li et al. (2016) report that the size of the expressed HIV reservoir and a patient’s drug regimen correlate with the time between ART suspension and viral rebound to detectable levels. We incorporate this information and viral rebound times to parametrize our model. The results we will discuss represent first early towards a model that can make predictions of a patient’s rebound time distribution based on patient characteristics, and help identify patients with expected long viral rebound delays. We will further discuss epidemiological implications of treatment suspension. ART is invaluable in preventing onwards transmission by controlling infection; similarly, individuals with controlled infection post-ART will have low risk of transmission. But that risk will increase at viral rebound, at which time an individual would re-initiate ART. We will also discuss model predictions that can be used to guide management of treatment suspension.