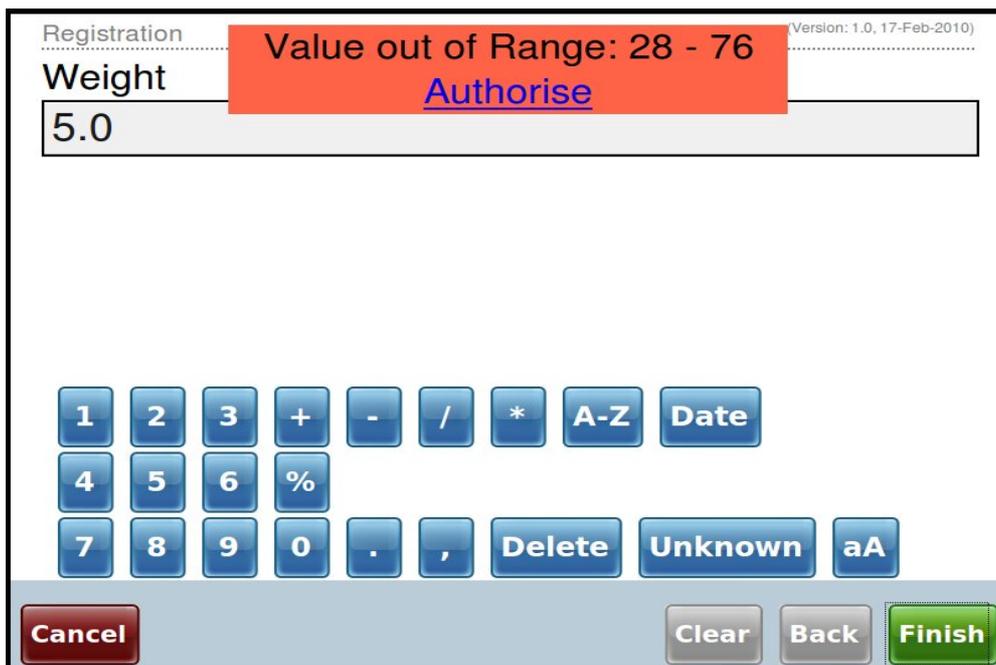


## Dynamic and Two-Level Data Validation

The ability to perform data validation during the time of entry can significantly improve data quality. Traditional approaches to validation use range checking. Upper and lower limits are set and data outside the range is disallowed.

Increased accuracy can be achieved using dynamic range checking. Here, the upper and lower limits are not fixed. Rather the limits move dynamically based on a rule or set of rules. For example, in Figure S10 below you can see a “Value out of Range” warning indicating the valid range for the patient’s weight is between 28 and 76 Kg. This range is calculated using the patient’s date of birth and a look-up table of weights for upper and lower limits for a patient of a specific age to the closest year.



The screenshot shows a software interface for patient registration. At the top, it says "Registration" and "Version: 1.0, 17-Feb-2010". Below this, the label "Weight" is followed by a red error message box that reads "Value out of Range: 28 - 76" and a blue link "Authorise". The text "5.0" is entered in the input field below the label. At the bottom of the screen, there is a numeric keypad with buttons for digits 1-9, 0, a decimal point, a comma, and symbols for plus, minus, division, and multiplication. There are also buttons for "A-Z", "Date", "Delete", "Unknown", and "aA". At the very bottom, there are buttons for "Cancel", "Clear", "Back", and "Finish".

Figure S10: Screen showing out of range error message

A major limitation of this approach to range checking is that if it is too constrained then the system may actually prevent the user from entering an extreme but valid data value. To get around this problem we have implemented two-level range checking. Here we create two ranges. The first range traps values in the traditional way, preventing unfeasible values from being entered. The second range traps values that are unlikely but plausible if they were

valid outliers. This technique slows the user down, forcing them to look closely at what they have entered. If the user recognizes they have made a transposition error for example, typing 27 instead of 72, then they can enter the correct value. However, if the value is truly 27, then the user can authorize the unlikely but correct value by entering their password, allowing the data to be saved. We believe the process of slowing the user down and forcing them to take a second look at the data value improves data accuracy.