**Summary of data transformation steps and coding of variables**

1. Transcribed all data excluding all names into 'The Hollies Spreadsheet 1'

2. Removed all addresses & transcribed into 'The Hollies numeric data 1934-38' and 'The Hollies numeric data 1948-60'

3. Changing categorical data into numerical data in            "                    "                    "

   (a) **Readmission (Excel column B)**

   | No readmission | 1 |
   |---|---|
   | Readmitted the same day: 'observation → definite TB | 2 |
   | Readmitted having been to Seacroft with another infection | 3 |
   | Readmitted having been to another hospital or health institution | 4 |
   | Readmitted for observation having been discharged well but deteriorated | 5 |
   | Readmitted for a second period of observation | 6 |
   | Readmitted for a third period of observation | 7 |
   | Readmitted for a fourth period of observation | 8 |

   (b) **Gender: (Excel column C)** Girl = 1   Boy = 2
       (NB this is the opposite to the common convention of labelling boys as 1 and girls as 2)

   (c) **'same child as' - children entered on the register twice. (Excel column E)**
   This is quite common, either because a child admitted for observation is found to have TB and is re-entered on the register, or the child is transferred to an infectious disease hospital and then readmitted to The Hollies after recovering from diphtheria, scarlet fever etc. One child was readmitted for observation, having been discharged. In the last 4-5 years there were many readmissions, commonly of sibling groups. Identifying children admitted more than once was relatively easy because Excel remembers previous addresses so, in spreadsheet 1, once an address starts to be typed in, a previous identical address appears in the cell.  If children have moved address between admissions it is possible that a few readmissions will have been missed. The index number and year of admission are listed on each occasion

   (d) **Diagnostic code (Excel Column F)**

   | Pulmonary TB (including pleural)(including RA1 and RA2 in 1950s classification) | 1 |
   |---|---|
   | TB glands | 2 |
   | Pulmonary TB + glands (or other non-pulmonary site) | 3 |
   | 'Observation' also 'TBCon' from 1955 onwards | 4 |
   | Bone and joint TB | 5 |
   | Pulmonary + Bone/joint TB | 6 |
   | TB peritonitis | 7 |
   | Bone and abdomen 8 | 8 |
   | Miliary | 9 |
   | TB meningitis | 10 |
   | 'Education' from 1956 onwards | 11 |
   | 'Welfare' from 1958 onwards (few of these) | 12 |
   | 'Children left in hall with note' - a group of 3 siblings – taken to a children's home next day | 13 |

   (e) **Sibling** (of someone else admitted to The Hollies) **(Excel Column L)** Yes=1 No=2

(f) **Outcome code (Excel Column M)**

| | |
|---|---|
| Pulmonary TB improved | 1 |
| Pulmonary TB arrested | 2 |
| Non-pulmonary improved | 3 |
| Non-pulmonary arrested | 4 |
| Not improved | 13 |
| **Outcomes for pre-tuberculous / observation** | |
| Negative | 5 |
| Improved | 6 |
| Definite PTB (relating to a child admitted for observation | 7 |
| Still observing | 11 |
| Definite abdominal glands | 12 |
| Taken home too soon before outcome evaluated | 8 |
| Transferred to another hospital before | 9 |
| Child ran away (no parental involvement recorded) | 10 |
| Discharged home for lack of discipline | 14 |
| Died | 15 |

(g) **Transfer codes (Excel Column N)**

| | |
|---|---|
| To home, or another relative, as planned | 1 |
| Parents took home before planned discharge | 2 |
| Transfer to Seacroft (always with an infectious disease) | 3 |
| Diagnosed with TB and 'readmitted' immediately | 4 |
| Transferred to hospital for surgery | 5 |
| Transferred to other hospital (Killingbeck or LGI or St James's) | 6 |
| Transferred to a children's home | 7 |

(h) **Will be readmitted (Excel Column O)**
This uses the same coding as (a) but is a means of showing what will happen to this child in the future.
For example, a child transferred to Seacroft will be readmitted from Seacroft or a child discharged home will be readmitted at a later date. It is a means of ensuring that children readmitted for whatever reason can be analysed as individual children and not double-counted

(i) **Not Leeds. (Excel Column P)**
Address data is **not** copied over from the original spreadsheet to the numeric data only spreadsheet.
However, in the 1950s, a few children start being admitted from outside Leeds. This is worth noting so an extra column has been added for Leeds = 1 and Not Leeds =2

(j) **Year (Excel Column Q)**
The year is listed at the beginning of each year group by admission date. This is included simply to help orientate a reader of the database.
N.B. when the data are analysed, year groups are defined by the date when the child was discharged which is the usual convention in longitudinal health service analyses.

D Cundall (last updated 24.1.24)