Dataset Description and Variable List

***Indefatigable* Training Ship Register Books Volumes 1-11 (1865-1912)**

Authors

Dr Eric Schneider (LSE), Principal investigator, e.b.schneider@lse.ac.uk
Dr Pei Gao (now NYU Shanghai), Research officer, p.gao@nyu.edu

Sources

National Museums Liverpool, Maritime Archives and Library, *Indefatigable* and National Sea Training School for Boys, Register Books vol. 12-40, D/IND/3/2/1-11.

Transcription Process

The dataset was transcribed from digital images of the microfiche of the original records held in the Merseyside Maritime Museum Library and Archives in Liverpool. The transcription took place between April and August 2015 and was conducted by a record transcription company that had experience transcribing historical, handwritten sources. Each record was transcribed by two people with discrepancies being checked by a supervisor. Any questionable entries after that were flagged to the PI who provided guidance on transcription. The quality of the digitised microfiche was low at times, so as a part of the checks (described in more detail below), the PI and RO went through and checked entries that could not be transcribed (especially for anthropometric measures and dates) against the original documents in Liverpool.

A sample page of the document has been provided to give future researchers an idea of what the original document looked like. The record is for boy 06009, John Edward Booth born in 1879 and admitted to the *Indefatigable* in 1891. Not all variables were transcribed because they were not systematic enough to provide useable data. Variables not transcribed included complexion, hair colour, eye colour, religion, parents' character, and boy's conduct on board and after leaving the ship.

The PI and RO also performed a series of checks on the transcribed data to ensure that the transcription was accurate. This included checking to make sure that the day, month and year transcriptions were plausible: i.e. they were in the correct range (1-12 for months, 1-31 for days and 1850-1994 for years) and they followed the correct ordinal pattern (birth date, then admission date, then discharge date). We also checked the anthropometric measurements in a couple of cases: 1) if the values were out of the correct range (0-11 inches, 0-13 lbs when stone were used) 2) if the discharge height was lower than the admission height; 3) if the measurement was implausibly high or low; 4) if the computed WHO Z-score was implausibly high or low, we checked the anthropometric measures and the birth and admission dates. Many odd cases were verified as correct in the original. When implausible results (mainly declines in height between admission and discharge) were verified, we left the measurements intact in the data but coded the anthro_meas_err (anthropometric measurement error) variable

as equal to 1. We excluded these from our analysis of the data, but they have been deposited here.

Variables

Variable naming and labelling principles

Different prefixes to the data refer to time periods or people that the variable relates to.

- b_ denotes variables related to the boy's birth
- ad_ denotes variables related to the boy's admission to the *Indefatigable*
- dis_ denotes variables related to the boy's discharge from the *Indefatigable*
- f_ denotes variables related to the boy's father
- m_ denotes variables related to the boy's mother
- p_ denotes variables related to the boy's parents
- rel_ denotes variables related to the boy's relatives (if mother and father are not present)

Different suffixes also indicate different treatments of the data

- _m indicates that the variable was calculated using the imputed age for boys whose exact birthdate (to the month) was not given. For these boys, we assumed their age to be their age at last birthday plus 0.5. The implications of imprecise ages on anthropometric measures is discussed at length in appendix B of Gao and Schneider (2018).

Whenever verbatim is listed in the variable label, this indicates that this variable has not been cleaned or streamlined and reflects the original document precisely.

Variable list

id	Unique id for each boy in the dataset. The first 2 digits refer to the volume in the original archival record and the last 3 digits are the page in the record where the child's data was recorded (each child had their own 2-page spread). This should ease further verification of the data in the future.
forename	Forename of boy (verbatim)
surname	Surname of boy (verbatim)
male	All children in the *Indefatigable* were boys. Male = 1
b_day	Day of birth (if present in original document – no imputation)
b_month	Month of birth (if present in original document – no imputation)
b_year	Year of birth (if present in original document – no imputation)
b_date	Stata formatted birth date (if day, month and year present in original)
b_cohort_m	Real or imputed birth year of boy. For a subset of the later data (and the majority of the early data), precise birth dates were not recorded in the

	records: only age at last birthday was given (see ad_age_last below). For these, we have assumed the child's age is the age at last birthday plus 0.5 (ad_ageyears_m) and have calculated the birth year by subtracting the imputed age from the admission date.
b_location	Birth location of boy (verbatim)
b_county	Birth county of boy (verbatim)
b_district_1851	Registration district in which boy was born (1851 boundaries)
b_county_1851	Registration county in which boy was born (1851 boundaries)
b_latitude	Latitude of birth location. This is not a precise measurement based on an address. We have just located the place name listed in b_location. It was only used to place individuals into 1851 registration districts and counties.
b_longitude	Longitude of birth location. See notes on b_latititude
ad_day	Admission day
ad_month	Admission month
ad_year	Admission year
ad_date	Stata formatted admission date
ad_age_last	Age at last birthday at admission (if present in original document – no imputation)
ad_agemonths_m	Real or imputed age at admission in months. If an precise birth date is available (i.e. a month and year of birth or age in years and months is given), then this is the exact age at admission in months calculated from subtracting the birth date from the admission date. If only the age at last birthday is available, then this is the age at last birthday plus six months. The implications of the imputed ages for anthropometric measures is discussed at length in Appendix B of Gao and Schneider (2018).
ad_ageyears_m	Real or imputed age at admission in years. If an exact age is available (i.e. a day, month and year of birth is given), then this is the exact age at admission in years calculated from subtracting the birth date from the admission date. If only the age at last birthday is available, then this is the age at last birthday plus 0.5 years. The implications of the imputed ages for anthropometric measures is discussed at length in Appendix B of Gao and Schneider (2018).
ad_heightft	Height at admission (ft)
ad_heightin	Height at admission (in)
ad_heightcm	Total height at admission (cm)
ad_weightkg	Weight at admission (kg)
ad_bmi	BMI at admission (kg/m^2)

ad_zwfa_m	Weight-for-age z-score at admission relative to the 2007 WHO growth reference. ad_agemonths_m was used as the age measure for calculating the z-scores.
ad_zhfa_m	Height-for-age z-score at admission relative to the 2007 WHO growth reference. ad_agemonths_m was used as the age measure for calculating the z-scores.
ad_zbfa_m	BMI-for-age z-score at admission relative to the 2007 WHO growth reference. ad_agemonths_m was used as the age measure for calculating the z-scores.
ad_read	Reading score at admission to the ship (verbatim). The tests and scale that were used to measure the boys changed dramatically over time, so great care should be taken in interpreting this variable.
ad_write	Writing score at admission to the ship (verbatim). The tests and scale that were used to measure the boys changed dramatically over time, so great care should be taken in interpreting this variable.
ad_school	School that boy attended before joining the *Indefatigable* (verbatim)
ad_employment	Work that the boy was doing before admission to the ship (verbatim)
ad_latitude	Latitude of boy's location before entering the *Indefatigable*. When the boy had been living with his parents or other relatives before admission, this is based on the parent or relative's address. When the boy entered the ship from a poor law union or workhouse as listed in the wherefrom variable, the location refers to the workhouse of the poor law union which we found in Higginbotham (2012). When a precise address could not be found, we used context from other variables such as ad_school and wherefrom to assign a location in a local area.
ad_longitude	Longitude of boy's location before entering the *Indefatigable*. See notes under ad_latitude for more information on construction.
ad_district_1851	Registration district in which boy was living before admission (1851 boundaries)
ad_county_1851	Registration county in which boy was living before admission (1851 boundaries)
f_forename	Father's forename (verbatim)
f_surname	Father's surname (verbatim)
f_desert	= 1 if the father deserted the child; = 0 if the father lived with the child prior to admission or died before admission, determined by the presence of other information about the father (occupation or address); = "." (blank) if there is no information listed about the father
f_dead	= 1 if the father is dead;

	= 0 if the father lived with the child prior to admission or deserted the child, determined by the presence of other information about the father (occupation or address); = "." (blank) if there is no information listed about the father
f_occ	Father's occupation (verbatim)
f_occ_hisco	HISCO code associated with father's occupation according to van Leeuwen, Maas and Miles (2002)
f_occ_hiscostatus	HISCO status code for father's occupation derived from van Leeuwen, Maas and Miles (2002, pp. 315-319) necessary for shifting from HISCO to HISCLASS
f_occ_hisclass	HISCLASS classification of father's occupation according to van Leeuwen and Maas (2011).
f_occ_hiscam	HISCAM classification of father's occupation according to Lambert *et al.* (2013)
m_forename	Mother's forename (verbatim)
m_surname	Mother's surname (verbatim)
m_desert	= 1 if the mother deserted the child; = 0 if the mother lived with the child prior to admission or died before admission, determined by the presence of other information about the mother (occupation or address); = "." (blank) if there is no information listed about the mother
m_dead	= 1 if the mother is dead; = 0 if the mother lived with the child prior to admission or deserted the child, determined by the presence of other information about the mother (occupation or address); = "." (blank) if there is no information listed about the mother
m_occ	Mother's occupation (verbatim)
m_occ_hisco	HISCO code associated with mother's occupation according to van Leeuwen, Maas and Miles (2002)
m_occ_hiscostatus	HISCO status code for mother's occupation derived from van Leeuwen, Maas and Miles (2002, pp. 315-319) necessary for shifting from HISCO to HISCLASS
m_occ_hisclass	HISCLASS classification of mother's occupation according to van Leeuwen and Maas (2011).
m_occ_hiscam	HISCAM classification of mother's occupation according to Lambert *et al.* (2013)
p_address	Parent's address (verbatim)
wherefrom	This is the location of the child before entering the ship (verbatim). For children living with their parents, this tends to be the city or town of residence of the parents. However, in some cases the children were not living with their

	parents, and this variable reflects where they came from, for instance from a workhouse, poor law union, or orphanage.
rel_forename	Forename of first listed relative. Relative variables are generally only recorded if the boy is not living with his parents.
rel_surname	Surname of first listed relative
rel_forename2	Forename of second listed relative
rel_surname2	Surname of second listed relative
rel_relationship	Relationship of relative to child (verbatim)
rel_address	Relative's address (verbatim)
dis_day	Discharge day
dis_month	Discharge month
dis_year	Discharge year
dis_date	Stata formatted discharge date
dis_agemonths_m	Real or imputed age at discharge in months. Follows same pattern as ad_agemonths_m.
dis_ageyears_m	Real or imputed age at discharge in years. Follows same pattern as ad_ageyears_m.
dis_heightft	Height at discharge (ft)
dis_heightin	Height at discharge (in)
dis_heightcm	Total height at discharge (cm)
dis_weightkg	Weight at discharge (kg)
dis_bmi	BMI at discharge (kg/m^2)
dis_zwfa_m	Weight-for-age z-score at discharge relative to the 2007 WHO growth reference. dis_agemonths_m was used as the age measure for calculating the z-scores.
dis_zhfa_m	Height-for-age z-score at discharge relative to the 2007 WHO growth reference. dis_agemonths_m was used as the age measure for calculating the z-scores.
dis_zbfa_m	BMI-for-age z-score at discharge relative to the 2007 WHO growth reference. dis_agemonths_m was used as the age measure for calculating the z-scores.
dis_read	Reading score at discharge from the ship (verbatim). The tests and scale that were used to measure the boys changed dramatically over time, so great care should be taken in interpreting this variable.

dis_write	Writing score at discharge from the ship (verbatim). The tests and scale that were used to measure the boys changed dramatically over time, so great care should be taken in interpreting this variable.
dis_occ	Occupation of child after leaving the *Indefatigable* (verbatim). Often reports the ship, captain, length of term and location that each boy was apprenticed with after leaving the *Indefatigable*. If the boy did not continue on as a sailor other information is listed.
dis_salary	Salary of boy after leaving the *Indefatigable* (verbatim). This record was transcribed verbatim and at times it was not clear whether the earnings were per month or week or whether the amounts referred to shillings or pence.
anthro_meas_err	This variable is equal to 1 if there is a verified height or weight measurement in the original record which has been deemed implausible. Most of these are cases where height has decreased between admission and discharge, but there are also a few very implausible outliers.

References

Gao, P., & Schneider, E. B. (2018). The Growth Pattern of British Children, 1850-1975. *LSE Economic History Working Paper*, forthcoming.

Higginbotham, P. (2012). The Workhouse Encyclopedia. Stroud: The History Press.

Lambert, P. S., Zijdeman, R. L., van Leeuwen, M. H. D., Maas, I., & Prandy, K. (2013). The Construction of HISCAM: A Stratification Scale Based on Social Interactions for Historical Comparative Research. *Historical Methods: a Journal of Quantitative and Interdisciplinary History*, *46*(2), 77–89. http://doi.org/10.1080/01615440.2012.715569

Van Leeuwen, M., Maas, I., & Miles, A. (2002). HISCO: Historical international standard classification of occupations. Leuven: Leuven University Press.

Van Leeuwen, M., & Maas, I. (2011). HISCLASS: A historical international social class scheme. Leuven: Leuven University Press.

Schneider, E. B. (2018). Sample Selection Biases and the Historical Growth Pattern of Children. *LSE Economic History Working Paper*, 1–31.