# Jörg Baten and Thomas Keywood, Univ., Tübingen

# Individual data sets of human stature: a global sample of samples

# Introduction

Height research became popular after the ground-breaking work by Robert Fogel in the 1970s, starting a largescale project which was taken over by Richard Steckel and John Komlos in the 1980s (Fogel 1994, Steckel and Floud 1997, Komlos 1989). Recently, height research has been expanded to cover developing countries and give broad global overviews (Baten and Blum 2014). Height is a function of net nutrition which is determined by the quality of gross nutrition and the prevailing disease environment. At the individual level, genetic variability is important, but looking at several hundred height measurements for whole countries or regions, the influences of nutrition and health are usually stronger than any potential genetic changes.

The datasets presented here cover a wide variety of countries and world regions. Providing them in a comparable and publicly accessible format will have important advantages for future research: Firstly, it will allow comparisons with new evidence that scholars might find in archives. Secondly, providing these public access data sets will enable researchers who are new to the field of anthropometric study to more thoroughly understand the challenges of this kind of analysis. Thirdly, the comparability of these data sets (perhaps in combination with additional regional or occupational variables) will allow researchers to answer new questions.

In the following text, we briefly describe the data sets and mention some background information. For example, the study of some samples has to take into account potential selectivity by social group, by labour market development and by region, if the focus is on understanding macroeconomic developments. While this was always a topic of anthropometric analysis, the rigidity of bias analysis varied a bit – this is fully understandable, as the sub-discipline developed new methods to do this (Baten, 2015). Hence, in the following, we will briefly discuss the individual datasets, and whether they are potentially influenced by selectivity issues or not. Various types of samples with their special selectivity risk have been distinguished: Conscript data, which reflects the whole male population of a certain age, has a reputation for not being selective at all (although it is possible that some conscripted recruits were still in the last phases of growth, potentially resulting in downwardly biased measurements for conscript heights, but this is a different issue; see Baten, 2000b). Similarly, anthropological samples cover usually the whole population of geographic unit (but potential regional biases needs to be studied if macroeconomic developments are of interest). Other types of samples are more at risk of selectivity issues, such as volunteer army recruits and prison samples, which might differ from the underlying population by (1) the level of height and also by (2) changes in height over time, if measurements were taken continuously over various years. If height was only recorded in one or in very few years, then changes in selectivity due to labour market developments are less likely (Baten 2015). As these issues have to be taken into account, they are documented in the following. The aim of this collection of datasets was to provide the datasets in a relatively standardized format, naming the variables in a way that allows comparisons. Some datasets have additional variables that are not included here, otherwise the samples provided here would have been very heterogeneous. We always list a typical study that allows the reader to learn more about each dataset.

We provide data containing most of the following variables (depending on availability):

age age

bdec birth decade

breg birth regions

rreg residence regions

byear year of birth

crime if crime is included in the data set (for prison samples)

ethn ethnic goup (if reported)

female 1, if female (otherwise male)

ht\_cm height in cm (if height is given in centimetres)

ht\_foot height in feet (if height is given in feet and inches), or ht\_varas if height was in

 varas

ht\_inch height in inches (if height is given in inches, or in feet and inches)

ht\_line height in lines (if height is given in feet, inches and lines)

original\_measure measure in which the height was originally taken

No\_ number (if original dataset has a running number)

occ occupation or occupational group

occ\_par occupation of parents

qyear year, in which height was recorded

bco country of birth (if distinguished)

rco country of residence (if distinguished)

regional\_intermediate name of regional unit in breg or rreg (if region available)

religion religion (if available)

**East Asia**

China: Prisoners in Australia

Literature/Source: Morgan, S. L. 2009. Stature and economic development in South China, 1810–1880. *Explorations in Economic History*, 46(1): 53 – 69.

Dataset: \_\_ea\_cn\_1810\_priso\_2009\_morgan\_prisoners.dta

Stephen Morgan (2009) collected a dataset on Chinese prisoners in Australia (among many other datasets) who were mostly convicted for a number of small crimes, with relatively few of them convicted for major crimes. As Morgan explains in his article, this dataset is potentially subject to two different selectivity biases, one of them being migrant selection: The Chinese migrants who went to Australia might have differed from the remaining population, plus the degree and type of selectivity might have changed over time. A large push of immigration from China to Australia occurred during the Australian Gold Rush of the 19th century, as Morgan (2009) explains, and most migrants came from the Southeastern Guangdong province. The second potential selectivity issue can occur when the individuals analysed were imprisoned. Poorer (and often shorter) people could be expected to be more likely to be imprisoned (although in reality, this is not always true, perhaps because tall stature is an asset in physical criminal activities, see Baten and Blum, 2012). There might also be changes in prisoner selectivity over time (for example, during economic boom phases, the opportunity costs of criminal activity are higher, and vice versa in crisis periods). In order to approach these selectivity issues, Morgan compared his source on Chinese migrants in prison with other datasets, and in other studies; non-migrants in a large sample of railway employees, and non-criminal migrants, for example. These have been analysed jointly, for instance, by Baten, Ma, Morgan and Wang (2010), who indicate that the resulting trends are relatively similar across datasets, even after coming from very distinct institutional contexts and different recruitment periods. The Chinese migrant prison dataset contains a large time dimension, spanning almost the entire 19th century. The regional focus is clearly on southern China and the province of Guangdong in particular. Occupations are recorded in the dataset, with labourers and miners mentioned most frequently. The dataset only consists of men, since it comes from prison inmates in jails that only imprison male criminals.

# China: Hospital Dataset from Guangdong

Whyte, G.D. 1911. Notes on the Height and Weight of the Hoklo People of the Kwangtung Province, South China. *The Journal of the Royal Anthropological Institute of Great Britain and Ireland*, 41(1): 278–300.

Dataset: \_\_ea\_cn\_1870\_hospi\_1911\_Whyte\_Yuxuan\_ed\_Kwangtung.dta

The hospital dataset from Guangdong (China) is based on examinations in the hospital of an English Presbyterian Mission in a fishing town south of Swatow. The birth years covered range from 1829 to the 1890s.

This institutional context probably provides no major selectivity issues as a wide coverage of illnesses is given (i.e., not only severe ones), and sampling is not restricted to sick patients: healthy hospital interns, employees, patients’ relatives and hospital visitors were also included. Only males older than seventeen have been measured, although further analysis might focus on age groups 20-50 in order to avoid age-specific height biases (Baten and Blum 2012).

The measurements were taken in English inches and no occupational data was available. The regional information for all observations is Guangdong, formerly named Kwangtung or Canton Province (on the South China Sea).

**Eastern and East-Central Europe**

Yugoslavia: Macedonian Anthropological Sample

Miszkiewicz, B., & Schade, H. 1961. Anthropologische Struktur der mazedonischen Bevölkerung. *Materiały i prace antropologiczne*, 62(1): 5–71.

Dataset: \_\_ee\_mk\_1920\_anthr\_1961\_Miszkiewicz.dta

The Macedonian dataset stems from an anthropological survey conducted in 1956, which sought to detect differences in height between the genders and across three of Macedonia’s ethnic groups. The survey includes observations from cities and town across the country and includes individuals between the ages of 21 and 50 (birth decades between 1900 and 1930). The dataset contains 307 individuals of the Aromun, Muslim and “Macedonian” ethnic groups. Due to the anthropological strategy to sample in a representative way, no obvious social or labour market selectivity issues arise. The original measurements were taken in centimetres and no occupations were listed.

Austria-Hungary: Habsburg Military Dataset

Komlos, J. 1985. Stature and Nutrition in the Habsburg Monarchy: The Standard of Living and Economic Development in the Eighteenth Century. *The American Historical Review*, 90(5): 1149–1161.

Komlos, J. 1989. Nutrition and Economic Development in the Eighteenth Century Habsburg Monarchy. An Anthropometric History. Princeton, NJ: Princeton University Press.

Dataset: \_\_ee\_eeu\_1720\_milit\_1885\_komlos\_incl\_at\_hu\_ua\_cz\_sk.dta

Modern countries included:

Austria

Czech Republic (“Moravia”, “Bohemia” in breg)

Hungary

Ukraine/Poland (“Galicia”)

Komlos collected a very large sample on soldiers in the Habsburg Empire who were born during the 18th and early 19th centuries. The observations from the 18th century are primarily characterised by volunteer recruits, so the trend has to be scrutinised for potential selectivity. Komlos (1985, 1989) did this by comparing the results obtained from volunteer data to those from various alternative datasets from other institutional contexts and other years of height measurement to make sure that the observed trends were not caused by the specific institutional and labour market contexts of the sample. The data set covers observations on Hungary, Lower Austria, Galicia, Bohemia and Moravia (=today’s Czech Republic). The occupational coding is done by the large occupational groups. The original measurements were performed in Austrian inches, birth decades run from 1720s – 1840s.

# Austria-Hungary: 18th century Military Boarding Schools

Komlos, J. 1986. Patterns of children's growth in East-central Europe in the eighteenth century. *Annals of Human Biology*, 13(1): 33–48.

Dataset: \_\_ee\_eeu\_1760\_schoo\_2009\_komlos\_incl\_at\_hu\_ua\_cz\_sk\_children.dta

Modern countries included:

Austria (“Steiermark”, “Upper Austria”, “Lower Austria”, “Carinthia”, “Tyrol”)

Croatia

Czech Republic (“Moravia”, “Bohemia”, “Silesia” in breg)

France

Germany (“Germany”, “Saxony”)

Hungary

Italy

Romania (“Transylvania”)

Ukraine/Poland (“Galicia”)

The 1760s Eastern European school dataset covers a number of Eastern European countries (see Komlos 1986). It is a dataset based on data from military boarding schools. The type of selectivity present is not absolutely clear, but in general, the parents of the students were probably more often in the military, government activities, and other middle and upper class occupations. The data considers birth years from the 1730s to the 1810s (Komlos 1986).

The measurements were taken in Austrian feet and inches. There is no occupational data available. The regional information was collected for all observations in 16 regions: Moravia, which is the eastern part of today’s Czech Republic; Bohemia, which lies in the western Czech Republic; Silesia, which is today mainly in Poland as well as the northern Czech Republic and eastern Germany; Steiermark, which is in today’s south-eastern Austria; Hungary; Upper Austria; Lower Austria; Carinthia, which lies in Austria, Italy and Germany; Saxony, in eastern Germany; France; Tyrol, which lies partly in today’s northern Italy and partly in today’s western Austria; Transylvania, which is in today’s east-central Romania; Croatia and a further “other” category.

Czech Republic: Prison Datasets in Pancrac and Repy

Stegl, M. & Baten, J. 2009. Tall and Shrinking Muslims, Short and Growing Europeans: The Long-Run Welfare Development of the Middle East, 1850-1980. *Explorations in Economic History*, 46(1): 132–148.

Dataset: \_\_ee\_cz\_1850\_priso\_2009\_stegl\_baten\_pankrac.dta

Dataset: \_\_ee\_cz\_1850\_priso\_2009\_stegl\_baten\_repy.dta

The two datasets on prisons in Pancrac and Repy – located in the Austro-Hungarian Empire during the 19th century, and in the territory of today’s Czech Republic – were prisons for males and females respectively. The usual selectivity problems associated with height data from prisons need to be considered. The range of survey years is very small for Pancrac, hence there was probably not much change in labour market conditions between survey years. Records were only taken between 1902 and 1904, and all three years were relatively similar in terms of economic prosperity, particularly as they were all after the minor crisis of 1901. Occupations are reported in German, the birth places in a mixture of German and Czech, and all measures were taken in centimetres. The birth years from Pancrac range from the 1850s to the 1880s, although there are only 431 observations recorded over this period. The dataset was originally collected by Katerina Hodinova (2007), but was not published by her – she included in it in her bachelor thesis. Stegl and Baten (2009) used these height records in a comparative study of European and Middle Eastern samples. The same applies to the Repy dataset, which is a data set of 338 female prisoners.

Austria: Military Conscripts of the late 19th Century

Schlesiger, C. 2016. Regionale Unterschiede von Lebensstandard und Humankapital in Österreich. Unpublished BA thesis, Univ. Tuebingen.

Dataset: \_\_ee\_at\_1860\_consc\_2010\_schlesinger\_ba.dta

A modestly sized conscript dataset on Austria was collected by Christian Schlesiger (2016). It contains height measurements between 1878 and 1911. The ages of the conscripts were between 20 and 22 years. The birth years covered range from 1857 to 1890. Austrian conscription covered the whole population with very few exceptions at that time. The geographic coverage includes parts of lower Austria as well as Styria and Carinthia. Although occupations were recorded in the original lists, they did not follow a consistent convention. Since Schlesiger did not record the occupations in large numbers, we did not include them in the public access dataset. As the heights of the recruits were recorded before the military selection process, selectivity is most likely not a strong issue in this dataset.

**Latin America/Caribbean**

Argentina: Military Census 1927

Baten, J., Pelger, I. & Twrdek, L. 2009. The anthropometric history of Argentina, Brazil and Peru during the 19th and early 20th century. *Economics and Human Biology*, 7(3): 319 – 333.

Dataset: \_\_la\_ar\_1870\_censu\_2009\_baten\_pelger\_twrdek.dta

The dataset on Argentina is a quite unusual dataset because it records the entire Argentinian population without social selectivity biases. The sample taken covers the birth years from the 1870s to the 1910s, and constitutes a relatively representative sample of provinces and cities from all over Argentina (Baten, Pelger and Twrdek 2009 provide a map). The original data is stored at the general archive of the army in Buenos Aires. The dataset was collected by Linda Twrdek. The importance of this dataset also stems from the fact that Argentina had an interesting stagnation in height which was visible in other samples taken previously, but these samples came from prisons or military records after a selection process. This representative (male) dataset allowed for comparisons (and actually confirms) the trends estimated on the basis of prison and other military samples. Apart from geographical information, the dataset also lists the occupations of the surveyed individuals. The original measurements were conducted in centimetres. The variation by birth cohort comes from the fact that all relevant ages were included, from 17 to 53 years. Due to the relatively high nutritional status in Argentina, it seems likely that most of the height growth was completed by the age of 17. The highest age group included was 53 years old; which is not yet meaningfully affected by shrinking.

Cuba: Army Recruits

Twrdek, L. 2010. Cuba, the ‘always most faithful island1’: Welfare Implications from Spanish Colonial Rule to North American Dependency (1870-1910)Cuba. Dissertation available at www.ub.uni-tuebingen.de

Dataset: \_\_la\_cu\_1870\_milit\_2010\_twrdek.dta

The data on Cuba comes from Cuban army registers and measures the heights of recruits between 1902 and 1951. Linda Twrdek (2010) collected the dataset from the national archive of Cuba. The earliest third of the data was collected in feet. Twrdek found out that the English foot measure was used for recording heights in this early phase, thereafter centimetres are used. There is a minimum height requirement applied to enlisted recruits, hence truncated regression is necessary. The minimum height requirement was 162.56 cm for the early period, until 1913. The selection process into the army beyond the minimum height requirement are not perfectly clear, although Twrdek found that there is evidence that the recruits share characteristics of the general population, as represented in national censuses; speaking for broad representativeness. For example, the ethnic composition was similar. Moreover, the skill levels of the occupations recorded between the two groups are broadly similar, though only some occupations were systematically compared to census data. Regional selectivity is less of a problem, as both provinces of western and eastern Cuba are included. Occupations in the dataset are not individual occupations, but Armstrong skill categories, which have been developed for 19th century censuses (this classification is explained in Twrdek 2010). The birth decades covered with a reasonable number of cases are the 1870s to the 1900s.

Peruvian Prisoner Dataset

Twrdek, L., & Manzel, K. 2010. The seed of abundance and misery: Peruvian living standards from the early republican period to the end of the guano era (1820–1880). *Economics & Human Biology*, 8(2): 145–152.

Dataset: \_\_la\_pe\_1820\_priso\_2010\_twrdek\_manzel.dta

The dataset by Twrdek and Manzel (2010) is a dataset which was created for two prisons, in Lima and Guadalupe, between 1866 and 1909. It contains evidence on prisoners who committed common crimes such as fraud and theft, but also, very prominently, homicide and other forms of violence. It covers the birth decades between the 1820s and 1880s. A large share of foreigners was included (28%) and while there are many more male prisoners (3,889), there are also 503 female prisoners among the total of 4392 prisoners. The regional focus is on the capital and the coastal zones of Peru, although 20 percent of the prisoners come from the other regions. Typical prison selectivity needs to be studied, but the authors have already compared the occupational composition of the prisoners with one of the Peruvian national occupational censuses. They find that farmers, for example, are underrepresented and that black individuals are over-represented by far. For Peru, it is more difficult to compare selectivity over time, as no heights from other sources have been mobilized so far.

United States: West Point Cadet Dataset

Komlos, J. 1987. The height and weight of West Point cadets: dietary change in antebellum America. *The Journal of Economic History*, 47(4): 897–927.

Dataset: \_\_na\_us\_1820\_schoo\_1987\_komlos\_westpoint.dta

The West Point cadets’ dataset comprises height data for over 4,000 applicants, taken upon their medical examinations at West Point Military Academy. Although problems of selectivity may arise due to the self-selection of the applicants, their economic backgrounds are relatively diverse (Komlos 1987). However, no black applicants appear in the sample and all applicants were male. The regional scope of the dataset covers 46 US states of birth and includes birth decades between 1820 and 1880. All cadets were between the ages of 15 and 26 at the time of each survey, meaning that some were still growing at the time of their medical examinations. The original measurements were taken in British inches.

United States (South Carolina): Citadel Military College Dataset

Coclanis, P.A., Komlos, J. 1995. Nutrition and Economic Development in Post-Reconstruction South Carolina: An Anthropometric Approach. *Social Science History*, 19-1, 91-115.

Dataset: \_\_na\_us\_1860\_milit\_2017\_coclanis\_komlos\_citadel.dta

The Citadel Military College dataset comprises height data for 6,534 applicants, taken upon their medical examinations. Although problems of selectivity may arise due to the self-selection of the applicants, their economic backgrounds are relatively diverse (Coclanis and Komlos 1995). The regional scope of the dataset covers many US states, but more than half are from South Carolina. It includes birth decades between 1860s and 1930s. All cadets were between the ages of 14 and 26 at the time of each survey, meaning that some were still growing at the time of their medical examinations. The original measurements were taken in British inches.

United States (Maryland) Dataset of free African-Americans

Komlos, J. 1992. Toward an anthropometric history of African-Americans: the case of the free blacks in antebellum Maryland. In *Strategic factors in nineteenth century American economic history: A volume to honor Robert W. Fogel*. University of Chicago Press: 297–329.

Dataset: \_\_na\_us\_1750\_freeb\_1992\_komlos.dta

This dataset was compiled from the Maryland Certificates of Freedom, which were produced between 1806 and 1864 for individuals born between 1742 and 1851. The aim of these certificates was to provide legal proof that a black person was free and not a runaway slave. Consequently, Komlos (1992) studied aspects of selectivity related to which groups of slaves received freedom (or were born free). The dataset contains nearly 15,000 observations, of which more than half were female. It also contains data on the county of birth (within Maryland) for each individual, their skin tone, and whether or not they were born a slave. No age restrictions were applied, meaning that the dataset includes individuals between the ages of 1 and 80, so some were still growing or had already undergone shrinking at the age of measurement. Original measurements were taken in British inches.

United States (Georgia) Convicts Dataset

Komlos, J., & Coclanis, P. 1997. On the puzzling cycle in the biological standard of living: the case of antebellum Georgia. *Explorations in Economic History*, 34(4): 433–459.

Dataset: \_\_na\_us\_1800\_priso\_1997\_komlos\_georgia.dta

This dataset includes height data for convicts from Georgia State Penitentiary, Milledgeville, collected between 1817 and 1885. The place of birth (state, country or world region – e.g. Europe) of each convict is listed, as well as the county in which the conviction took place. The dataset also includes the occupation of each prisoner, and which crime they were sentenced for. Since this is a prisoner dataset, we have to be aware of the usual potential prison selectivity issues, although Komlos and Coclanis (1997) also discuss these by comparing the data to other samples from other institutional backgrounds. The original measurements were taken in British inches. The dataset includes both black and white Georgian prisoners and foreign-born convicts. Convicts not born in the “American South” make up 7.8% of all observations (3.4% were European-born, 3.1% from the Northern USA, and 1.3% from the rest of the world).

**South Asia**

Indian Anthropological Dataset

Sources:

Ganguly, P., 1976. Physical Anthropology of the Nicobarese; Anthropological Survey of India. Calcutta: Government of India.

von Ujfalvy, E. K., 1884. Aus dem westlichen Himalaya. Erlebnisse und Forschungen. Leipzig, F.A. Brockhaus.

Waddell, M. B. 1908. The Tribes of the Brahmaputra Valley; Journal of the Asiatic Society of Bengal 69-3.

Dataset: \_\_sa\_in\_1860\_anthr\_div.dta

The dataset on India combines three different anthropological studies, one on the north-west, one on the north-east and one on the Andaman and Nicobar Islands in the south-east. The dataset covers two earlier studies, which mostly represent birth decades of the 1850s to the 1900s, and then a larger dataset on the early 20th century. Social selectivity is not an issue as these are anthropological measurements, but this dataset is not regionally representative for all of India. There has been some debate about whether early anthropologists measured very “typical” physical characteristics of the ethnic groups which they studied, implying selectivity. This can be counterchecked by comparing the early samples with the migration evidence provided by Brennan et al. (1995) and other publication by this team of authors, which has not yet been made publicly available at an individual level.

Pakistan and adjacent Indian regions: Anthropological Datasets

Sources:

Mazumdar, S. K., 1976. A Biometric Study on the Tribes of North-Western Himalayan Region; Anthropological Survey of India, Calcutta: Ghosh.

von Ujfalvy, E.K., 1884. Aus dem westlichen Himalaya. Erlebnisse und Forschungen. Leipzig, F.A. Brockhaus.

Dataset: \_\_sa\_pk\_1840\_anthr\_div.dta

The dataset on Pakistan and adjacent Indian regions combines two anthropological studies, an earlier one from the 1880s and a later one from the 1970s. Both document heights in the Kashmir and north-western Himalaya region. Social selectivity is not present (anthropological sampling), but there is strong regional selectivity due to concentrating in the north-east of Pakistan and the north-west of India. The periods covered in this dataset are the birth decades from the 1840s to the 1900s. The latter dataset is small.

**Southeast Asia**

Myanmarese (Burmese) Anthropometric Dataset

Bassino, J. P., & Coclanis, P. A. 2008. Economic transformation and biological welfare in colonial Burma: Regional differentiation in the evolution of average height. *Economics & Human Biology*, 6(2): 212–227.

Dataset: \_\_se\_mm\_1860\_anthr\_2008\_bassino\_coclanis.dta

Bassino and Coclanis (2008) employ a dataset of heights in Myanmar between the 1810s and the 1880s in their study of colonial administration and economic development. The dataset is based on a British colonial ethnographic survey conducted in 1904. This dataset covers men between the ages of 21 and 85. The original measurements were made in British inches and the survey was carried out in order to discern whether the average statures of 13 distinct ethnic groups differed across six districts in each of Upper- and Lower Burma. Social selectivity is not a problem here. Occupations were not recorded in this survey.

Philippine Anthropological Datasets

Sources:

Bean, R.B., 1909. Filipino Ears-A Classification of ear types; The Philippine Journal of Science IV-1 (Jan. 1909).

Bean, R.B. 1909. II. Filipino Types: Found in Malecon Morgue; The Philippine Journal of Science IV-1 (Jan. 1909)

Bean, R.B. 1909. III. Filipino Types: Racial Anatomy in Taytay; The Philippine Journal of Science IV-1 (Jan. 1909)

Bean, R.B., Planta, F.S., 1911. The Men of Cainta; The Philippine Journal of Science, D.VI-1 (1911)

Schebesta, P., 1956. Die Negrito Asiens, Band I, St. Gabriel Verlag, Wien.

Wastl, J., 1957. Beitraege zur Anthropologie der Negrito von Ost-Luzon, Anthropos 52.

Weisbach, A., 1878. Koerpermessungen verschiedener Menschenrassen, Verlag von Wiegandt, Hemppel & Parey, Berlin, 1878.

Dataset: \_\_se\_ph\_1860\_anthr\_div.dta

The dataset on the Philippines is a combination of seven different anthropological studies which do not suffer from social or labour market selectivity and cover a number of different regions of the Philippines. The “Negrito” ethnic groups are documented particularly well. The birth decades cover the period from the 1860s to the 1920s.

**Sub-Saharan Africa**

# Various African countries: Liberated slaves in Cuba

# Eltis, D. & Nwokeji, G. U. 2002. Characteristics of Captives Leaving the Cameroons for the Americas, 1822-37. *The Journal of African History*, 43(2): 191 – 210.

Dataset: \_\_af\_afr\_1820\_slave\_1982\_eltis\_havana\_paris\_foot.dta

Modern countries included:

Angola

Benin

Gongo-Brazzaville

Cameroon

Guinea-Bissau

Nigeria (-East, -West)

Sierra Leone

This dataset was created by David Eltis (2002) and covers the heights of liberated slaves born in various African countries who were freed (and documented) in the Cuban capital of Havana. The dataset is available online, on the website slavevoyages.org, jointly within an even larger dataset on liberated slaves in Sierra Leone (and some smaller places, such as Saint Helena and the Bahamas). We have chosen the Havana dataset because the age statements were quite likely done by the slaves themselves, allowing insights into their potential selectivity by studying age-rounding and hence numeracy (Eltis and Nwokeji 2002, Baten and Cappelli 2016). The dataset contains the most likely country or group of countries where the slaves came from based on their ports of origin; this was suggested by David Eltis as a rough identification possibility, although there is clearly a measurement issue with the slaves who were transported over further distances to each port, as he mentions. David Eltis and his colleagues also undertook large efforts to identify the ethnic origins of these individuals by their names; sometimes the African ethnic group was mentioned, but in a less systematic way, hence these are not covered in the dataset provided here. The period roughly covers the birth decades from the 1780s to the 1830s. Selectivity issues are carefully discussed in the paper by Eltis and Nwokeji (2002). The conclusion is that several positive and negative selectivity issues about height were not very substantial, and cancelled each other out. The most commonly used measure in Cuba during this period, for this kind of purpose, was the Paris foot measure according to Baten and Blum (2014).

Kenyan Colonial Military Dataset

Moradi, A. 2009. Towards an objective account of nutrition and health in colonial Kenya: A study of stature in African army recruits and civilians, 1880–1980. *The Journal of Economic History*, 69(3): 719–754.

Dataset: \_\_af\_ke\_1890\_milit\_2009\_moradi.dta

Moradi’s military height dataset from Kenya is derived from military records from the headquarters of the Kenyan Armed Forces in Nairobi. It records 1,845 recruits with birth decades between 1890 and 1940, aged between 16 and 52. The dataset also includes the place of birth for each soldier, their religion and the years in which they were enlisted. The dataset includes some young soldiers with the potential to grow further. Further, selectivity must be considered when dealing with military data, as soldiers are usually taller than the rest of the population, on average. Moradi (2009) analysed these issues carefully and concluded that selectivity is much less of an issue for the cohorts recruited during WWII, when 20% of the wage labour force was part of the armed forces (whereas during peacetime, recruitment was limited and focused on the so-called “martial races”, i.e., native tribes with a greater reputation for combat, and otherwise low incomes). The original measurements were taken in British inches.

**Western/Central Europe**

# Germany: 1730s Military Dataset (Bavaria, Palatinate, Upper Palatinate)

Baten, J. 1997. *Ernährung und wirtschaftliche Entwicklung in Bayern, 1730-1880*. Doctoral thesis, University of Munich, Munich. Original in German, English translation available from author.

Dataset: \_\_we\_de\_1730\_milit\_1997\_baten\_bavaria\_palatinate\_upperpal\_.dta

The German 1730s military dataset is a dataset based on military lists created by the Kingdom of Bavaria and its predecessors. The period studied ranges from the 1720s to the 1770s.

The data is partially made up of volunteer military recruits, which comes with their own special selectivity issues, and partially from a drafting system which is described in the study by Baten (1997). There is also a minimum height requirement for recruits, which requires truncated regression analysis.

The measurement was taken in Bavarian feet or Rhenish feet, depending on which (sub-) principality recruited the soldiers. Occupations are only mentioned for about 1,500 soldiers out of a total of over 18,000. This is partially because only a fraction of the soldiers were recorded with occupations and partially because farming was not listed as an occupation – this is especially relevant as farming was the most widespread occupation (80% of Bavaria’s economy was agricultural at this time), and second and third sons of farmers who did not inherit, joined the army in particularly large numbers.

Regional information was collected for all observations in the three broad regions – Bavaria, the southern half of today’s Bavaria; Upper Palatinate, which is the central-eastern quarter of Bavaria today; and Palatinate, which is the eastern half of Rhineland-Pfalz as well as northwestern Baden-Württemberg today.

# South/Southwest Germany: Military Conscripts

Baten, J. 2000a. Economic Development and the Distribution of Nutritional Resources in Bavaria, 1797-1839: An Anthropometric Study. *Journal of Income Distribution*, 9(1): 89–106.

Dataset: \_\_we\_de\_1800\_consc\_2000\_baten\_tcft.dta

Joerg Baten collected a dataset on conscripts in the Kingdom of Bavaria (Southern Germany) which also included the region of Palatinate (today the eastern part of today’s German federal state of Rhineland Pfalz, Southwest Germany). The dataset is not socially selective, because all heights of the male 20.5-year-olds were measured before the decision to recruit a candidate soldier was taken (by drawing lots). Only very few exceptions have to be taken into account, such as priests and migrants who had left the country before the obligatory military medical examination took place. Regionally, all parts of Bavaria are represented (though Southern Bavaria is oversampled and should perhaps obtain smaller weights, if the focus would be on representativeness for Bavaria). The dataset lists the occupations of the candidate’s parents in addition to their own occupation. The parental occupation is even more informative for height research, as height is mostly determined in the first years of life. Occupations are recorded in German, ages are always 20.5, and the first observations were recorded in Bavarian inches before later being changed to centimetres. The decades covered by this dataset are the 1810s to the 1840s, and with 21,000 observations, the dataset is quite large.

# West Germany: Prussian Military Conscripts

Baten, J. & Fertig, G. 2009. Did the Railway Increase Inequality? A Micro-Regional Analysis of Heights in the Hinterland of the Booming Ruhr Area During the Late 19th Century. *The Journal of European Economic History*, 38(2): 263–299.

Dataset: \_\_we\_de\_1880\_consc\_2010\_baten\_fertig\_arnsberg.dta

This dataset covers military conscripts from the Prussian army of northwestern Germany and, in particular, the Sauerland region to the south of the Ruhr area. It was collected in order to study the development of a rural, agricultural region next to the rapidly industrializing Ruhr district. There are no selectivity issues present because recruitment officials inspected all males of this age group. Regionally, a broad mix of districts is available, making the data set representative for this region of Germany. The dataset covers the second half of the 19th century or, to be more precise, two years of the late 19th century as well as a limited number of observations in between.

# South/Southwest Germany: Prison Inmates from Kaisheim and Wasserburg

Baten, J. 1997. *Ernährung und wirtschaftliche Entwicklung in Bayern, 1730-1880*. Doctoral thesis, University of Munich, Munich. Original in German, English translation available from author.

Dataset: \_\_we\_de\_1830\_priso\_1997\_baten\_kaisr.dta

Dataset: \_\_we\_de\_1810\_priso\_1997\_baten\_wasserbg.dta

The two prison datasets Kaisheim and Wasserburg cover male and female prison inmates in southern Germany. The usual selectivity problems of prison records should be taken into account. However, Baten (1997) intensively discusses how selective-on-observables the prison records were by comparing the share of occupations with the census distribution of these occupations, and also finds that, while covering a less skilled share of the population, quite a variety of different occupations are included. The trends in height are also similar to (non-selective) military conscript data covering the same region of Germany (see the discussion of the Bavarian conscripts, Baten 2000a and 2000b), which suggests that selection on unobservables does not play a substantial role. The regional differentiation is by Bavarian districts, “Regierungsbezirk”, of which about 300 existed in Bavaria, as well as some foreign prisoners. The dataset covers the intermediate period of the 19th century, from the 1810s to the 1870s.

# French Military Dataset

Komlos, J., Hau, M. & Bourguinat, N. 2003. An anthropometric history of early-modern France. *European Review of Economic History*, 7(2): 159–189.

Dataset: \_\_we\_fr\_1670\_milit\_2003\_komlos\_hau.dta

Michel Hau and Nicolas Bourguinat collected a dataset on the French army that covers a very early period, namely the 1670s up to the 1770s. The dataset is a classical volunteer army dataset. Komlos et al. (2003) discuss potential effects of supply and demand for recruits during various periods of boom and crisis (crisis enlistment of taller soldiers, relative to normal or boom periods), but they do not assign much weight to labour market selectivities. The – not strictly applied – minimum height requirement changed over time. It was, for example, 6.2 feet in 1763. The birth regions of French soldiers are quite comprehensive, hence there is no regional selectivity, but social selectivity and labour market selectivity are carefully studied by the authors. The regional disaggregation is by place (for cities, sometimes including surrounding rural areas) and governments and occupations are also included in the dataset.

# The Netherlands: Female Prison Dataset

de Beer, H. 2010. Physical stature and biological living standards of girls and young women in the Netherlands, born between 1815 and 1865. *The History of the Family*, 15(1): 60–75.

Dataset: \_\_we\_nl\_1820\_priso\_2010\_debeer.dta

The prison dataset from the Netherlands covers female prisoners and is therefore particularly interesting. The usual prison selectivity problems apply here and have to be carefully studied, as was done by de Beer (2010) in his study of this prison. The regional and occupational breakdown of the dataset is as follows: The dataset covers birth decades over the period 1790 to 1880 and includes 629 places of birth across all of the Netherlands. There are also 69 distinct occupations listed in Dutch.

Norwegian Military Dataset

Wiegel, B. V., 2015. The Biological Standard of Living in Norway during the 18th Century and Thereafter: Does Selectivity Drive the Results? Unpublished BA. Thesis,. Univ. Tuebingen.

Dataset: \_\_we\_no\_1740\_milit\_2008\_wiegel\_ba.dta

The Norwegian dataset covers military recruits which have been collected by Wiegel (2015). They cover the birth decades of the first three quarters of the 18th century and are subject to the usual military sample biases. Occupational data is not available to a large extent, hence it has not been included in the dataset. The regional selection covers many different parts of Norway.

# Portuguese Military Datasets

Stolz, Y., Baten, J., & Reis, J., 2013. Portuguese living standards, 1720–1980, in European comparison: heights, income, and human capital. *The Economic History Review*, 66(2): 545–578.

Dataset: \_\_we\_pt\_1730\_consc\_2013\_stolz\_baten\_reis\_mhr.dta

The Portuguese dataset covers three different types of datasets. The most recent one is from a general conscription period of the 19th century. The data was collected by Jaime Reis and there are no substantial selectivity issues, contrary to what Bodenhorn et al. (2017) claim about this dataset. The earliest part is a classical military dataset with a volunteer soldier component. It has to be scrutinised for military selectivity as was done to a certain extent by Stolz et al. (2013). Finally, in the middle period of the 19th century Portugal had a combined system of recruitment. The dataset covers the whole period from the 1710s to 1910s. Occupations are included and the regional aggregation is both on the level of birth places and places of residence.

# UK Prisoners sent to New South Wales

Oxley, D. 2006. ‘Pitted but not pitied’ or, does smallpox make you small? *The Economic History Review*, 59(3): 617–635.

Dataset: \_\_we\_uk\_1780\_priso\_9999\_oxley\_nsw\_smallpox\_study\_also\_ie.dta

The UK dataset of prisoners who were transported to New South Wales, Australia, was collected by Deborah Oxley and her colleagues and analysed in order to answer several questions, including a study on the effects of smallpox. The dataset needs to be carefully studied in terms of potential selectivity, as Oxley (2006) did. The level of regional disaggregation include the larger geographic units of Ireland, London, as well as rural and urban England. The coverage of the data is from the 1750s to the 1820s.

UK: Prisoners from Victorian London at Wandsworth prison

Horrell, S., Meredith, D., & Oxley, D. 2009. Measuring misery: Body mass, ageing and gender inequality in Victorian London. *Explorations in Economic History*, 46(1): 93–119.

Dataset: \_\_we\_uk\_1800\_priso\_2009\_horell\_oxley\_wandsworth\_smallpox.dta

Horrell, Meredith and Oxley used prisoner data from London’s Wandsworth prison (and other sources) in their study of gender inequality during the Victorian Era. The time period involved includes birth decades from the 1780s to the 1850s. The potential prisoner data selectivity issues might apply here as well, although Horrell et al. (2009) studied this carefully. The original measurements were taken in British feet and inches, and occupations were grouped into five skill categories.

# References

Bassino, J. P., & Coclanis, P. A. 2008. Economic transformation and biological welfare in colonial Burma: Regional differentiation in the evolution of average height. *Economics & Human Biology*, 6(2): 212–227.

Baten, J. 1997. *Ernährung und wirtschaftliche Entwicklung in Bayern, 1730-1880*. Doctoral thesis, University of Munich, Munich. Original in German, English translation available from author.

Baten, J. 2000a. Economic Development and the Distribution of Nutritional Resources in Bavaria, 1797-1839: An Anthropometric Study. *Journal of Income Distribution*, 9(1): 89–106.

Baten, J. 2000b. Heights and real wages in the 18th and 19th centuries: an international Overview. *Jahrbuch fuer Wirtschaftsgeschichte* 2000-1, pp. 17-32.

Baten, J. 2015. Strategies to cope with potential labor market bias and other selectivities in height research. Working Paper Uni. Tuebingen.

Baten, J., & Blum, M., 2012. Growing Taller, but Unequal: Biological Well-Being in World Regions and Its Determinants, 1810-1989. *Economic History of Developing Regions* 27 (2012), pp. S66-S85. V.

Baten, J., & Blum, M., 2014. Why are you tall while others are short? Agricultural production and other proximate determinants of global heights”, *European Review of Economic History* 18, 144–165.

Baten, J. & Cappelli, G., 2016. The Evolution of Human Capital in Africa, 1730 – 1970: A Colonial Legacy? CEPR Working Paper 11273.

Baten, J. & Fertig, G. 2009. Did the Railway Increase Inequality? A Micro-Regional Analysis of Heights in the Hinterland of the Booming Ruhr Area During the Late 19th Century. *The Journal of European Economic History*, 38(2): 263–299.

Baten, Ma, Morgan and Wang (2010)

Baten, J., Pelger, I. & Twrdek, L. 2009. The anthropometric history of Argentina, Brazil and Peru during the 19th and early 20th century. *Economics and Human Biology*, 7(3): 319 – 333.

Bean, R.B., 1909. Filipino Ears-A Classification of ear types. The Philippine Journal of Science IV-1 (Jan. 1909).

Bean, R.B. 1909. II. Filipino Types: Found in Malecon Morgue. The Philippine Journal of Science IV-1 (Jan. 1909)

Bean, R.B. 1909. III. Filipino Types: Racial Anatomy in Taytay. The Philippine Journal of Science IV-1 (Jan. 1909)

Bean, R.B., Planta, F.S., 1911. The Men of Cainta. The Philippine Journal of Science, D.VI-1 (1911)

### Bodenhorn, H., Guinnane, T. W.; Mroz, T. A.; 2017. Sample-Selection Biases and the Industrialization Puzzle Journal of Economic History, March 2017, v. 77, iss. 1, pp. 171-207

Brennan, L. McDonald, J & Shlomowith, R. 1995. The Variation in Indian Height. *Man in India* 75-4, pp. 327-337.

Coclanis, P.A., Komlos, J. 1995. Nutrition and Economic Development in Post-Reeconstruction South Carolina: An Anthropometric Approach. *Social Science History*, 19-1, 91-115

de Beer, H. 2010. Physical stature and biological living standards of girls and young women in the Netherlands, born between 1815 and 1865. *The History of the Family*, 15(1): 60–75.

Eltis, D. & Nwokeji, G. U. 2002. Characteristics of Captives Leaving the Cameroons for the Americas, 1822-37. *The Journal of African History*, 43(2): 191 – 210.

Fogel, R. (1994) “Economic Growth, Population Theory, and Physiology: The Bearing of Long-Term Processes on the Making of Economic Policy” American Economic Review 84: 369-95.

Ganguly, P., 1976. Physical Anthropology of the Nicobarese; Anthropological Survey of India. Calcutta: Government of India.

Horrell, S., Meredith, D., & Oxley, D. 2009. Measuring misery: Body mass, ageing and gender inequality in Victorian London. *Explorations in Economic History*, 46(1): 93–119.

Komlos, J. 1985. Stature and Nutrition in the Habsburg Monarchy: The Standard of Living and Economic Development in the Eighteenth Century. *The American Historical Review*, 90(5): 1149–1161.

Komlos, J. 1986. Patterns of children's growth in East-central Europe in the eighteenth century. *Annals of Human Biology*, 13(1): 33–48.

Komlos, J. 1987. The height and weight of West Point cadets: dietary change in antebellum America. *The Journal of Economic History*, 47(4): 897–927.

Komlos, J. (1989) Nutrition and Economic Development in the Eighteenth-Century Habsburg Monarchy: An Anthropometric History. Princeton: Princeton University Press.

Komlos, J. 1992. Toward an anthropometric history of African-Americans: the case of the free blacks in antebellum Maryland. In *Strategic factors in nineteenth century American economic history: A volume to honor Robert W. Fogel*. University of Chicago Press: 297–329.

Komlos, J., & Coclanis, P. 1997. On the puzzling cycle in the biological standard of living: the case of antebellum Georgia. *Explorations in Economic History*, 34(4): 433–459.

Komlos, J., Hau, M. & Bourguinat, N. 2003. An anthropometric history of early-modern France. *European Review of Economic History*, 7(2): 159–189.

Mazumdar, S. K., 1976. A Biometric Study on the Tribes of North-Western Himalayan Region; Anthropological Survey of India, Calcutta: Ghosh.

Miszkiewicz, B., & Schade, H. 1961. Anthropologische Struktur der mazedonischen Bevölkerung. *Materiały i prace antropologiczne*, 62(1): 5–71.

Moradi, A. 2009. Towards an objective account of nutrition and health in colonial Kenya: A study of stature in African army recruits and civilians, 1880–1980. *The Journal of Economic History*, 69(3): 719–754.

Morgan, S. L. 2009. Stature and economic development in South China, 1810–1880. *Explorations in Economic History*, 46(1): 53 – 69.

Oxley, D. 2006. ‘Pitted but not pitied’ or, does smallpox make you small? *The Economic History Review*, 59(3): 617–635.

Schebesta, P., 1956. Die Negrito Asiens, Band I, St. Gabriel Verlag, Wien.

von Ujfalvy, E.K., 1884. Aus dem westlichen Himalaya. Erlebnisse und Forschungen. Schlesiger, C. 2016. Regionale Unterschiede von Lebensstandard und Humankapital in Österreich. Unpublished B.A. thesis Univ. Tuebingen.

Steckel, R. H. & Floud, R. (eds.) (1997) Health and Welfare during Industrialization. Chicago: The University of Chicago Press.

Stegl, M. & Baten, J. 2009. Tall and Shrinking Muslims, Short and Growing Europeans: The Long-Run Welfare Development of the Middle East, 1850-1980. *Explorations in Economic History*, 46(1): 132–148.

Stolz, Y., Baten, J., & Reis, J. 2013. Portuguese living standards, 1720–1980, in European comparison: heights, income, and human capital. *The Economic History Review*, 66(2): 545–578.

Twrdek, L. 2010. Cuba, the ‘always most faithful island1’: Welfare Implications from Spanish Colonial Rule to North American Dependency (1870-1910)Cuba. Dissertation available at www.ub.uni-tuebingen.de

Twrdek, L., & Manzel, K. 2010. The seed of abundance and misery: Peruvian living standards from the early republican period to the end of the guano era (1820–1880). *Economics & Human Biology*, 8(2): 145–152.

von Ujfalvy, E. K., 1884. Aus dem westlichen Himalaya. Erlebnisse und Forschungen. Leipzig, F.A. Brockhaus.

Waddell, M. B. 1908. The Tribes of the Brahmaputra Valley; Journal of the Asiatic Society of Bengal 69-3.

Wastl, J., 1957. Beitraege zur Anthropologie der Negrito von Ost-Luzon, Anthropos 52.

Weisbach, A., 1878. Koerpermessungen verschiedener Menschenrassen, Verlag von Wiegandt, Hemppel & Parey, Berlin, 1878.

Wiegel, B. V., 2015. The Biological Standard of Living in Norway during the 18th Century and Thereafter: Does Selectivity Drive the Results? Unpublished BA thesis, Univ. Tuebingen.

Whyte, G.D. 1911. Notes on the Height and Weight of the Hoklo People of the Kwangtung Province, South China. *The Journal of the Royal Anthropological Institute of Great Britain and Ireland*, 41(1): 278–300.