

Constructing SHiP and an International Historical Coding System for Causes of Death

By Angélique Janssens

To cite this article: Janssens, A. (2021). Constructing SHiP and an International Historical Coding System for Causes of Death. *Historical Life Course Studies*, 10, 64-70. <https://doi.org/10.51964/hlcs9569>

HISTORICAL LIFE COURSE STUDIES

Not Like Everybody Else.
Essays in Honor of Kees Mandemakers

VOLUME 10, SPECIAL ISSUE 3
2021

GUEST EDITORS
Hilde Bras
Jan Kok
Richard L. Zijdeman



MISSION STATEMENT

HISTORICAL LIFE COURSE STUDIES

Historical Life Course Studies is the electronic journal of the *European Historical Population Samples Network* (EHPS-Net). The journal is the primary publishing outlet for research involved in the conversion of existing European and non-European large historical demographic databases into a common format, the Intermediate Data Structure, and for studies based on these databases. The journal publishes both methodological and substantive research articles.

Methodological Articles

This section includes methodological articles that describe all forms of data handling involving large historical databases, including extensive descriptions of new or existing databases, syntax, algorithms and extraction programs. Authors are encouraged to share their syntaxes, applications and other forms of software presented in their article, if pertinent, on the openjournals website.

Research articles

This section includes substantive articles reporting the results of comparative longitudinal studies that are demographic and historical in nature, and that are based on micro-data from large historical databases.

Historical Life Course Studies is a no-fee double-blind, peer-reviewed open-access journal supported by the European Science Foundation (ESF, <http://www.esf.org>), the Scientific Research Network of Historical Demography (FWO Flanders, <http://www.historicaldemography.be>) and the International Institute of Social History Amsterdam (IISH, <http://socialhistory.org/>). Manuscripts are reviewed by the editors, members of the editorial and scientific boards, and by external reviewers. All journal content is freely available on the internet at <https://openjournals.nl/index.php/hlcs>.

Co-Editors-In-Chief:

Paul Puschmann (Radboud University) & Luciana Quaranta (Lund University)
hislives@kuleuven.be

The European Science Foundation (ESF) provides a platform for its Member Organisations to advance science and explore new directions for research at the European level. Established in 1974 as an independent non-governmental organisation, the ESF currently serves 78 Member Organisations across 30 countries. EHPS-Net is an ESF Research Networking Programme.



The European Historical Population Samples Network (EHPS-net) brings together scholars to create a common format for databases containing non-aggregated information on persons, families and households. The aim is to form an integrated and joint interface between many European and non-European databases to stimulate comparative research on the micro-level.
Visit: <http://www.ehps-net.eu>.



Constructing SHiP and an International Historical Coding System for Causes of Death

Angélique Janssens

Radboud University Nijmegen & Maastricht University

ABSTRACT

SHiP is a network of European researchers studying mortality dynamics in port cities across Europe in the 19th and 20th centuries. All members make use of unique individual-level cause-of-death data for roughly the period 1850–1950 which allows the study of mortality to move beyond what was captured in nineteenth-century highly-aggregated national statistics. Apart from registering the individual cause of death, most datasets provide a wealth of information, such as name and address of the deceased, date of death, his/her age, sex, marital status, and religion and occupation of the deceased. Port cities are viewed as 'gateways of disease' in the same way that airports today function as hubs for the transmission of infectious diseases. The SHiP network aims to study the particular epidemiological profiles of the port cities in a truly comparative fashion across the different European maritime areas. To that end the SHiP team members have embarked upon the construction of a joint coding scheme, called ICD10h, which assigns codes to a large number of causes of death in a systematic way. Its main features are that the ICD10h coding scheme can deal well with large numbers of historical disease descriptions, from different linguistic areas in Europe, while at the same time it is able to connect to current day disease patterns.

Keywords: Causes of death, Coding scheme, Mortality, Port cities, Europe, 19th and 20th Centuries

e-ISSN: 2352-6343

DOI article: <https://doi.org/10.51964/hlcs9569>

The article can be downloaded from [here](#).

© 2021, Janssens

This open-access work is licensed under a [Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/), which permits use, reproduction & distribution in any medium for non-commercial purposes, provided the original author(s) and source are given credit. See <http://creativecommons.org/licenses/>.

1 SHiP: THE NETWORK

In 2017, the research network SHiP (Studying the history of Health in Port cities) was established.¹ SHiP is a network of scholars studying the dynamics of health and mortality change in port cities across Europe in the 19th and 20th centuries. The network makes use of individual-level cause-of-death data for the entire population of these cities for roughly the period 1850–1950. These datasets are truly unique as they enable us to go beyond what was captured in highly-aggregated national statistics which make use of extremely limited 19th-century disease classifications. In this way, we can evaluate health changes in depth, being able to study dis-aggregations, by individual disease, by age, sex, et cetera. We can thus reconstruct the epidemiological 'fingerprints' of European port cities and the way these changed in an exceptional period in the history of European health, in which life expectancy nearly doubled and infectious diseases sharply declined.

Researchers in the SHiP network regard port cities in our period of study as unique laboratories which enable us to better understand the long-term global evolution of mortality and health in dynamic environments. We assume that late 19th- and early 20th-century port cities were characterized by high rates of population turnover and by vast economic and industrial change. In that sense, port cities may be seen as pre-figurations of the highly mobile and dynamic societies that we live in today. However, port cities were not only hubs for the transmission of people and goods, but also acted as 'gateways of disease' in the same way that airports today function as hubs for the transmission of infectious diseases such as Covid-19, Ebola and Zika. Different types of travellers and migrants may have been very effective carriers and spreaders of microbes, sometimes by acting as their host and at other times by carrying goods which enabled microbes to travel.

From this it follows that port cities are also characterized by a particular epidemiological profile which is different from other cities. In general, port cities may have higher disease loads from infectious diseases, they may have suffered from a larger variety of infectious diseases, and epidemics may have arrived earlier in port cities than in other cities, resulting in higher casualties than elsewhere. Sea ports for instance played a major role in the global transmission of the third bubonic plague pandemic, which struck the world between 1894 and 1901 (Echenberg, 2007). Finally, port cities were not only connected across the sea with other ports, they also maintained important links to their immediate and wider hinterlands. Hence, changing health and mortality conditions in port cities may have had an important impact upon demographic and epidemiological change in Europe at large.

The SHiP network aims to study the particular epidemiological profiles of the port cities in a truly comparative fashion across the different European maritime areas (North Sea, Nordic Sea, Mediterranean) and the reconstruction of spatiotemporal patterns. In this way we wish to study the complex dynamics of demographic and epidemiological change and the relationships and interactions with large scale socio-economic change, migration flows, local and national health infrastructures, as well as differences resulting from diverging medical knowledge structures and other types of contextual factors. These ambitious aims imply the construction of joint methodological tools, the most important of which is a joint coding scheme which can be used to code historical causes of death in a strictly comparative fashion.

In this brief contribution I will present a discussion of what SHiP is about, and the comparative coding scheme we are designing within the network. In the following sections of the paper I will first discuss the port cities incorporated in the network, after which I will present the main features of the individual level data involved. I will conclude with a discussion of the international historical coding scheme that is being developed by a core team within the SHiP network.

2 SHiP: PORT CITIES

In the period covered by most of the datasets in the SHiP network world trade expanded vastly, and so did port cities. Whereas at the beginning of the century most countries contained large numbers

¹ See: <https://www.ru.nl/rich/our-research/research-groups/radboud-group-for-historical-demography-and-family/ship/>; or contact the author.

of towns which could be characterized as ports, by the second half of the century this was no longer the case. With the ever rising volume of quantities traded, only a few ports emerged that could handle these quantities and could meet the demands of the revolution in transportation. As Jürgen Osterhammel (2014) termed it, the 19th century is the golden age of *large* port cities. Sea ports in particular grew in significance; linking countries and continents, sea ports came to be the main transmission hubs in the world. The industrial revolution was key to the revolution in sea shipping and transportation. It facilitated the replacement of wood by iron and later by steel in the construction of ships, and the replacement of sails by steam. In the 20th century, further rapid technological change in sea shipping involved the introduction of oil-burning and motor-driven ships. Sea shipping continued to maintain its dominant position in overseas trade against promising newcomers such as air transport until the middle of the 20th century.

Large port cities in the period 1850–1950 were special worlds, harbouring cosmopolitan and dynamic populations with an enormous diversity in its labour market, on the one hand engaged in hard manual labour and on the other in highly skilled financial services or red-light districts. Major sea ports not only attracted foreigners from overseas, but were also magnets to individuals from surrounding areas looking for work or just waiting to begin their journey on one of the great ocean steam liners to the New World. Major port cities also developed into large urban and industrial centres, not just in Europe but also elsewhere. Overseas trade was an important driver of urbanization and industrialization: in 1850, 40% of cities with a population exceeding 100,000 inhabitants were ports (Osterhammel, 2014).

Map 1 *Map showing the European port cities incorporated in the SHIP network*



Source: Cartography Torsten Wiedemann, UGent

Map 1 shows the European port cities presently incorporated in the SHIP network. Clearly, large differences existed between them. Some of these port cities (Amsterdam, Antwerp, Copenhagen, Glasgow and Venice) had already passed the 100,000 inhabitants threshold in 1850, and grew out to even larger proportions after that, most notably Glasgow, Amsterdam and Copenhagen. Other ports

were smaller to begin with and experienced only modest population growth after 1850. Nevertheless these smaller ports acted as important transportation hubs in their own maritime areas. For instance, Palma was one of the most important seaports in the Mediterranean maritime area, also receiving steam shipping in the later decades of the 19th century. Also in other ways the SHiP ports display different features. Some cities had important administrative functions, for instance Copenhagen and Amsterdam which were national capitals, whilst other cities stand out for the highly industrialised profile, such as Glasgow. All cities were involved in trade and transport, but not all of them had shipyards or played a role in international migration routes to the Americas. However, we regard the differences between the port cities as excellent comparative research opportunities. Unfortunately, the coverage of cities in the different European maritime areas is somewhat unequal, in the sense that the northern and north-western part of the continent is rather well represented, but better coverage for the southern part of Europe and also the Baltic sea would be desirable.

3 SHiP: THE DATA

All members within the SHiP network make use of unique individual-level cause-of-death data for their cities. As was pointed out in the introduction, these data enable us to move beyond the highly aggregated mortality statistics at the national level and the limited 19th century disease classifications. Moreover, these national statistics are seldom available for the period prior to 1900 so that we cannot fully capture the decisive turning points in the historical decline of mortality. The common denominator of the datasets in the network is that in principle the data contain a single entry for each and every individual death in that particular city stating the deceased person's cause of death. The data however stem from different types of sources which vary from church registers, to civil death registers and separate civil cause of death registers. Sometimes cities combine more than one type of source, such as the Spanish port cities which have civil death registers, which include the cause of death, from 1871 onwards, but were preceded by church registers which also registered the causes of death. In the case of Amsterdam, special cause of death registers were initiated in 1853 by the city authorities in response to the increasing fear amongst the population of being buried alive. Hence, the city adopted the rule that all deaths needed to be verified by a doctor before burial could take place. The members of local medical society then pressed for a systematic verification of the cause of death to be reported before the death certificate could be issued. Hence from 1854 onwards the city was able to start a systematic registration of all deceased individuals together with the cause they died of (Neurdenburg, 1929). The resulting registers were kept until the end of 1940.

The temporal coverage of the different datasets varies to some extent. Some datasets start as early as the final decade of the eighteenth century (Rostock) or the early decades of the 19th century (Antwerp) whereas others begin as late as 1871 (Ipswich). However, apart from that latter town all other cities are well represented from the middle of the 19th century onwards reaching into the first half of the 20th century and sometimes even up until the 1970s (Glasgow).

Apart from registering the individual cause of death, most datasets provide a wealth of information, such as name and address of the deceased, date of death, his/her age, sex, marital status, and religion and occupation of the deceased. For some cities the data also contain info on the place of birth, the legitimacy of birth for infants, the names of the parents, and the name of the certifying doctor, and in some cases even the duration of the disease. The research potential is greatly increased when these data are linked to other sources, such as population registers, birth, death and marriage certificates, as well as sources containing cadastral information on housing and properties. In the case of the Amsterdam cause of death data, entries are currently being linked to life course data in the Historical Sample of the Netherlands and the cadastral data on rental values of the houses of the deceased. In addition, the increasing potential of GIS techniques and geospatial analytical techniques offer great opportunities to combine time and space in the analysis of causes of death and the mortality decline. Finally, the combination of socio-economic (address, occupation), demographic (age, sex, marital status) and disease data for all individuals opens up the possibility to clarify the *relative* impact of different mortality determinants, with unprecedented precision.

4 SHiP: THE CODING SCHEME

The wealth and the high level of detail of the cause of death data also poses challenges. The individual disease data are in most cases not only highly detailed and containing a larger number of different causes of death, also after standardisation of spelling varieties. However, we are also dealing with a long time span in which the meaning of medical terminology underwent fundamental changes, due to changes in diagnostic practices and the understanding of disease causation (Alter & Carmichael, 1999; Anderton & Leonard, 2015). There is furthermore reason to assume that medical understanding and registration traditions may have differed not only across time but also across places. Furthermore, it would be naive to assume that 19th century disease terms can be taken at face value. As Ann Hardy (1994) claimed, the historian wishing to use cause of death data is involved in 'an exercise in detective skills and a test of historical intuition (...) fraught with pitfalls'. Hardy was referring to the use of cause of death data pertaining to only one country, Great Britain which probably has the best record in the registration of historical causes of death. The attempt to study causes of death across space as well as time will increase the level of complexity. How to ensure that the European wide comparative and long term study of causes of death, which is considered so vital for a proper understanding of the mortality transition, can be conducted in a methodologically sound way?

To that end the SHiP team members have embarked upon the construction of a joint coding scheme, called ICD10h, which should operate much like the HISCO coding scheme for occupations across time and space. The ICD10h coding scheme assigns codes to a large number of causes of death in a systematic way. The ICD10h scheme is however not a classification scheme. A classification scheme places causes of death in a small number of groups of diseases for analytical purposes, whereas a coding scheme is nested within a classification scheme and can be used to be built up into different classification schemes. A coding scheme is therefore not tied to a particular classification scheme. The ICD10h system is adapted from a coding scheme used earlier by the Cambridge Group for the History of Population and Social Structure in a number of studies (see e.g. Reid, Garrett, Dibben, & Williamson, 2015). The most important features of the ICD10h coding scheme are that the system is flexible and detailed so that it can deal well with large numbers of historical disease descriptions, from different linguistic areas in Europe, while at the same time it is able to connect to current day disease patterns. This latter feature is sustained because the ICD10h is based upon the contemporary ICD10 system as designed by the World Health Organisation (the International Classification of Diseases, version 2016: www.who.int). A further important characteristic of the ICD10h coding scheme is that it avoids unwanted levels of interpretation of disease descriptions by coding 'words' rather than 'diseases'. In the paragraphs below this will be elaborated further.

The modern-day ICD10 system is based on alphanumerical 4-digit system organised in chapters and blocks of disease groups, the base version of which allows for as many as 14.000 different codes. Codes consist of one capital letter (ranging from A through to Z) followed by 2 digits, a dot and another digit (e.g. A00.1). Thus, for tuberculosis in different areas of the body the blocks numbered A15-A19 have been reserved, whilst the final digit following the dot indicates further specificities of the disease, such as type of confirmation of the disease. To create space for historical detail within each disease category the ICD10h system has added 2 additional digits to the right of the code (e.g. A00.100). Let us take as an example the different historical terms for typhoid fevers listed below and which may be found in 19th century sources. For each of these different terms a separate code is created within the block of typhoid fevers. This allocation is reached in two ways: based on the ICD10 system itself as well as on medical historical source material. The descriptions of enteric and gastric fever when entered into the ICD10 system take you to the block of typhoid fevers. For the historical terms bilious fever, colonial fever, and fog fever historical medical textbooks indicate that these terms were frequently used to indicate typhoid fevers. However, in order to track the different terminologies for typhoid fever over time and place it is important to keep them apart through the allocation of separate codes for each of these terms. This also allows us to trace the use of these older medical terms over time and between locations.

Table 1 *Historical terms for typhoid fever*

ICD10	ICD10h	Standardised CoD
A01.0 – Typhoid and paratyphoid fevers I Typhoid fever	A01.000	typhoid fever
	A01.001	enteric fever
	A01.002	bilious fever
	A01.003	gastric fever
	A01.004	colonial fever
	A01.005	fog fever

The principle to code 'words' and avoid unwanted or unnecessary interpretations of historical terminology may be further illustrated by way of an example frequently found in English sources, but also elsewhere in Europe, namely the disease terms 'teething' and 'dentition'. There is abundant historical proof that these terms were frequently used to refer to diarrheal disorders in infants, rather than any real dental problems. However, instead of allocating this term to a diarrheal block of codes, the ICD10h allocates it to a general code within the block for disorders of tooth development, namely K00.7. However, a special historical subdivision is created for the respective historical terms: 'teething' is coded as K00.700 and 'dentition' as K00.701. Researchers using the ICD10h coding system are then free when constructing their classifications to classify these terms as diarrheal disorders rather than as tooth development disorders.

The ICD10h database system will be expanded to include historical disease terms from all SHiP members and relating to all European countries or geographical areas represented in the network. To that end a dictionary of causes is being developed which shows the different linguistic varieties. However, different linguistic disease varieties may also refer to different disease meanings, hence these disease terms also need their own codes. The current ICD10h system for instance contains a special code for the Swedish disease term for stroke, namely 'slag', as this term is not only used for adults but also frequently for infants. It is thought to mean something else for infants than for adults. Hence, 'slag' is given its own code within the block of terms for stroke (I64), and accompanied by an age flag indicating that researchers need to be aware of age-specific meanings of that term. The justification for the creation of new historical codes not only requires evidence that these terms carried their own specific meanings at that time, but they also need to refer to more than just a handful of cases.

In October 2020 the SHiP coding scheme had its first official release to be used for a first round of testing within a group of SHiP members. This round of testing involves causes of death for infants only for nine different locations throughout Europe, resulting in nine case studies which all follow the same research questions and the same methodological approach and a comparable time frame from approximately mid-19th century to early decades of the 20th century. The aim is to determine and compare the dominant cause of death pattern of infants in different European port towns over time, and secondly, to test the use of the ICD10h coding system in practice. The results will be published as a special issue in the open access journal *Historical Life Course Studies*. The network envisages further rounds of testing by SHiP members, and eventually a public release to the wider scholarly community through a dedicated website. We also envisage the publication of a manual to guide researchers through the coding of their own cause of death data. At this moment we do not envisage the construction and publication of automatic coding routines.

REFERENCES

- Alter, G. C., & Carmichael, A. G. (1996). Classifying the dead: Toward a history of the registration of causes of death. *Journal of the History of Medicine and Allied Sciences*, 54(2), 114–132. doi: [10.1093/jhmas/54.2.114](https://doi.org/10.1093/jhmas/54.2.114)
- Anderton, D. L., & Leonard, S. H. (2004). Grammars of death: An analysis of nineteenth-century literal causes of death from the age of miasmas to germ theory. *Social Science History*, 28(1), 111–143. doi: [10.1017/S0145553200012761](https://doi.org/10.1017/S0145553200012761)

- Echenberg, M. (2007). *Plague ports. The global urban impact of Bubonic Plague, 1894–1901*. New York & London: New York University Press.
- Hardy, A. (1994). 'Death is the cure of all diseases': Using the general register office cause of death statistics for 1837–1920. *Social History of Medicine*, 7(3), 472–492. doi: [10.1093/shm/7.3.472](https://doi.org/10.1093/shm/7.3.472)
- Neurdenburg, M. G. (1929). *Doodsoorzaak en statistiek*. Amsterdam: H.J. Paris.
- Osterhammel, J. (2014). *The transformation of the world. A global history of the nineteenth century*. Princeton & Oxford: Princeton University Press.
- Reid, A., Garrett, E., Dibben, C., & Williamson, L. (2015). 'A confession of ignorance': Deaths from old age and deciphering cause-of-death statistics in Scotland, 1855–1949. *The History of the Family*, 20(3), 309–319. doi: [10.1080/1081602X.2014.1001768](https://doi.org/10.1080/1081602X.2014.1001768)