Routledge
Taylor & Francis Group

Population Studies
A Journal of Demography

ISSN: 0032-4728 (Print) 1477-4747 (Online) Journal homepage: https://www.tandfonline.com/loi/rpst20

Early-life mortality clustering in families: A literature review

Ingrid K. van Dijk

To cite this article: Ingrid K. van Dijk (2019) Early-life mortality clustering in families: A literature review, Population Studies, 73:1, 79-99, DOI: 10.1080/00324728.2018.1448434

To link to this article: https://doi.org/10.1080/00324728.2018.1448434

Published online: 04 May 2018.

Population Studies, 2019
Vol. 73, No. 1, 79–99, https://doi.org/10.1080/00324728.2018.1448434

Review Article

Early-life mortality clustering in families: A literature review

Ingrid K. van Dijk
Radboud University Nijmegen

Research on early-life mortality in contemporary and historical populations has shown that infant and child mortality tend to cluster in a limited number of high-mortality families, a phenomenon known as 'mortality clustering'. This paper is the first to review the literature on the role of the family in early-life mortality. Contemporary results, methodological and theoretical shortfalls, recent developments, and opportunities for future research are all discussed in this review. Four methodological approaches are distinguished: those based on sibling deaths, mother heterogeneity, thresholds, and excess deaths in populations. It has become clear from research to date that the death of an older child harms the survival chances of younger children in that family, and that fertility behaviour, earlier stillbirths, remarriages, and socio-economic status all explain mortality clustering to some extent.

Keywords: biodemography; child mortality; demographic methods; less developed countries; early-life mortality; family demography; historical demography; infant mortality; literature review; mortality clustering

[Submitted January 2017; Final version accepted January 2018]

Introduction

This paper offers a review of the causes of a phenomenon known as mortality clustering, defined in this study as the occurrence of excess infant and child mortality in a select group of families. Almost 25 years ago, Monica Das Gupta (1990, 1997) recognized that, despite high mortality rates in the Indian region that she was studying, infant and child mortality were concentrated in a subset of the families under study. Over recent decades, it has been shown that early-life mortality does indeed cluster in families, in both high- and low-mortality populations (Van Bodegom et al. 2012; Van Poppel et al. 2012), in historical populations (see, e.g., Edvinsson et al. 2005), and in present-day populations in the developing world (see, e.g., Zenger 1993; Omariba et al. 2008). Mortality clustering, or death clustering, has important implications for research and policy. For research methods, it implies that in determining the causes of early-life mortality, we should account for the correlated observations of children from the same family. For theories on early-life mortality, death clustering implies that we need to shift the focus from the child level to the level of the families in which children are embedded. Finally, for policy, it means that an infant or child death in a family may function as a red flag for further bereavement.

Over recent decades, many researchers have contributed to our understanding of the causes of mortality clustering in families. At the same time, these efforts have been fragmented: partially taking place in the field of (historical) demography and partially in development studies and anthropology, using different methodological approaches and therefore yielding insights into mortality clustering from varying analytical angles. Whereas some researchers have focused on the scarring effects of child deaths on the health of remaining siblings, others have analysed heterogeneity between families or mothers, analysed differences between high- and low-risk families, or assessed concentration in families in the

population. In this review paper, I attempt to place the approaches and results of these research efforts within a coherent framework. A systematic literature search was conducted using relevant search engines, including PubMed, Web of Science, and Google Scholar. Searching for literature on infant and child mortality or death clustering provided 681 titles to consider. I identified all papers discussing mortality clustering and removed those that did not consider infant or child mortality, overviews of the literature, contributions without data analyses (including research notes), presentations unaccompanied by papers, and earlier versions of selected works. This left 170 papers that were listed for further consideration. Full papers were carefully assessed to determine whether these contributions addressed differences in infant and child mortality between families or mothers: 53 papers met these criteria, while 117 did not. A further six papers and one book chapter were added from other sources. In total, 60 papers and chapters are included.

In this paper, I reflect on innovative research efforts and results, identify challenges, and explore avenues for future research that have opened up in part because new, large-scale data sets have become available, for both contemporary developing world populations and historical populations from the developed world. While there has been extensive research on spatial mortality clustering as well as mortality clustering in families, the focus of the current paper is the latter, and so research concerning spatial clustering is included only if the primary topic is early-life mortality clustering in families. To the knowledge of the author, this is the first attempt to review the literature on infant and child death clustering. The review is based on literature on mortality clustering in the first 60 months of life (i.e., both infant and under-five child mortality) in families; I refer to this collectively as 'early-life mortality clustering'.

Research on death clustering does not commonly deliver insights into the medical causes of death of individuals: instead, it aims to explore why some families are more vulnerable to disease and death among their children than others (for an exceptional case where medical causes are taken into account, see Lundevaller and Edvinsson 2012, on Rh disease). In historical and developing world populations, communicable diseases (such as diarrhoea and measles) constitute a major cause of death (Van Bodegom et al. 2012). Families have social, economic, and behavioural characteristics, which are risk factors for exposure and vulnerability to diseases (Mosley and Chen 1984; Sastry 1997). Death clustering seems to be caused by a complex interplay of all kinds of causes: not only do socio-economic characteristics seem important, but genes, the size of the family, maternal care practices, and the health of the mother also seem to be related to the risk of early-life mortality. But the risk factors are not exclusively the domain of the mother and other family members: socio-economic, genetic, behavioural, and environmental roots of individual susceptibility to death can be found at the child level and the community or population level too (Madise and Diamond 1995; Sastry 1997). These three analytical levels (child, family, and community) have been the focus in different strands of research within the mortality clustering literature, and all three are taken into account in this review.

The paper proceeds as follows. In the next section, the concept of mortality clustering is explored in depth. Four approaches to measuring and modelling death clustering are evaluated on their merits. Then I discuss current evidence, focusing in turn on the level of the child, the family, and the community. Finally, pressing challenges for future research are identified. Although extensive research has been conducted in some subfields and results have been reaffirmed by multiple scholars, other subfields have received more fragmentary attention, and yet others remain completely unexplored. First, I turn to the concept and measurement of early-life mortality clustering.

Early-life mortality clustering: meaning, measurement, and modelling

Mortality clustering is a crucial concept for understanding early death, as research has repeatedly shown that families play a pivotal role in the survival chances of infants and small children. In most studies on mortality clustering, the authors present the distribution of deaths across families in the population under study, using various measures. Often, the share of deaths in high-risk families is shown as a proportion of the total number of deaths in the population; in other cases, the proportion of mothers who experience a certain share of the deaths is reported. For instance, 12.6 per cent of mothers accounted for 62.2 per cent of child deaths in rural Punjab, India in the 1980s (Das Gupta 1990), while Edvinsson et al. (2005) found that in nineteenth-century Skellefteå, north-east Sweden, approximately 50 per cent of deaths were found in 10 per cent of the families. Typically, such numbers are used to show the extent to which clustering plays a role in mortality in the populations being studied.

Although sometimes presented as an indicator for mortality clustering, these numbers are not necessarily meaningful. In measuring whether populations experience inequality in the number of early deaths between families, it is essential to take both the size of the family and the binomial distribution of the probability of mortality into account. First, large families can be expected to lose more children than small families. If 62 per cent of the babies were born into 12 per cent of the families in rural Punjab at the end of the twentieth century, it would not be remarkable if these large families experienced 62 per cent of the total infant mortality. In that case, their mortality rate would be the same as that of the population as a whole. Second, next to family size we need to account for binomial distributions to assess whether mortality clusters in families in a population. If early deaths follow a binomial probability distribution, the implication is that not all families will experience the same realized mortality rate: there must be 'lucky' large families that escape early death among their children and 'unlucky' small families who have to bury multiple children. Even if one in five children dies before their fifth birthday, chance predicts that more than 10 per cent of large families with ten children will not lose a single child. Thus, in assessing whether deaths cluster in the population we should take both chance variations and the size of the family into account, and this constitutes a major methodological challenge.

A further challenge is related to the assumption of many statistical techniques commonly applied in infant and child mortality research that observations—in this case, observations on the mortality of siblings—are independent from each other. Considering that mortality clusters in high-risk families, observations on children originating from the same family are not independent (Beise and Voland 2002; Omariba et al. 2007). Thus, we have to account for these linked survival chances as well as chance variations and family size in our statistical models.

Many methodological solutions have been proposed for these issues, resulting in varying definitions of mortality clustering at three analytical levels: the children, their families, and their communities. Four different approaches are commonly used in the literature on mortality clustering in families. These approaches are summarized in Table 1. A first approach, used at the child level, is to use sibling death as an indicator of mortality clustering in sequence models. The second and third approaches are applied at the family level. In these cases, some authors use random intercept models assessing heterogeneity in the likelihood of death of a child at the level of the family as an indication of mortality clustering. Other authors distinguish high-risk from low-risk families by a set threshold, for instance, at least two deaths in the sibling group (Das Gupta 1990). The fourth and final approach uses the difference between the expected and observed numbers of families with a certain number of deaths at the community or population level as an indication of mortality clustering.

These variations in methods and definitions of mortality clustering have resulted in a rather heterogeneous field of research, split up by the analytical level of interest, in which results are not easy to compare and are seldom considered in an integrated fashion. In this paper, I attempt to reunite findings from these varying angles, exploring methods and findings, and identifying gaps in the literature. First, the four approaches are discussed in more detail.

The first approach models the impact of the death of one sibling on their other siblings, interpreting the relationship as a causal effect, using sequence models or dynamic models (Table 1, approach 1). Examples include the work by Curtis et al. (1993), Zenger (1993), Arulampalam and Bhalotra (2006, 2008), and Omariba et al. (2008). Analytically, the focus lies on the effect of the death of a sibling on survival of the other siblings after accounting for the correlated death risk of the sibling set. Neither parity nor the binomial chance distribution are usually taken into account, meaning that the focus is not on *excess* mortality, but on early-life death itself, as a result of a previous sibling's death. Furthermore, authors often use Markov models to assess scarring effects. These models assume a strict sequence of events: the death of an older sibling influences the death of a younger sibling. This assumption is not usually met in models that incorporate mortality after the first birthday and, consequently, there has been more attention on scarring effects on infants (<12 months) than on older children (12–60 months); for an exception, see Ikamari (2000). Finally, neither family nor community effects are incorporated in the models.

In the second approach, random intercept models are applied to focus on mother heterogeneity in the likelihood of death and survival of children, modelling family-level variables next to child-level variables (Table 1, approach 2). Often, these models include a description of the (explained) variance at the family level and sometimes the community level. Results commonly show the extent to which variation exists between families before and after

Table 1 Overview of methodological approaches to mortality clustering

Approach	Definition of mortality clustering	Example methods	Takes binomial chance distribution into account?	Insight into child-level variation?	Insight into family-level variation?	Insight into community-level variation?	Example studies
1	Sibling death	Sequence analysis (dynamic models)	No	Yes	No	No	Zenger (1993) Arulampalam and Bhalotra (2008) Omariba et al. (2008)
2	Mother heterogeneity	Variance of random intercept, usually in survival analysis (frailty models)	No	To some extent	Yes	Depending on approach	Sastry (1997) Janssens and Pelzer (2012) Scalone et al. (2017)
3	Families experiencing mortality beyond set threshold	Logistic regression	Depending on approach	No	Yes	No	Das Gupta (1990, 1997) Edvinsson et al. (2005) Vandezande and Matthijs (2013)
4	Difference between observed and expected numbers of deaths	Simulation models; binomial models	Yes	Depending on approach	Yes	No	Ronsmans (1995) Holmberg and Broström (2012) Lundevaller and Edvinsson (2012)

controlling for family- and child-level characteristics, and the variance of the mother- or family-level effects is interpreted as an indication of mortality clustering (for a discussion of the interpretation of model parameters, see Scalone et al. 2017). The models do not compare the observed distributions of deaths with the expected distributions, although solutions involving simulation techniques have been suggested (Holmberg and Broström 2012). The effects of unobserved family characteristics are assumed to be independent from other predictors in the model. Furthermore, the models assume that all children are affected equally by their family's characteristics and, therefore, offer limited insights into differences within families. Examples include work by Guo (1993), Sastry (1997), Bolstad and Manda (2001), and Janssens and Pelzer (2012).

A third approach also focuses on the family level, applying a threshold approach to distinguish between high- and low-risk families and using logistic regression analysis to model explanatory characteristics at the family level (Table 1, approach 3). While earlier authors used criteria such as sibling death (Das Gupta 1990) or multiple child loss (Das Gupta 1997; Kuate-Defo and Diallo 2002), more recent authors have used the size of the family and population mortality rates to determine which families experience mortality beyond what can be expected. For instance, Edvinsson et al. (2005) considered families to be high risk if their infant death rate was twice the population-level infant mortality rate or higher. While this approach to mortality clustering is intuitive and takes fertility and mortality patterns in the population into account, it should also be noted that there is an operationalization effect. In large families, it is more likely that at least one infant or child death occurs, and therefore, the likelihood of a large family being a low-mortality family is lower. This effect has been illustrated graphically by Vandezande (2012), who showed that small families were more likely to be classed as low-mortality families than large families. Furthermore, although this approach puts family characteristics in the spotlight, there is less opportunity to assess whether child or community characteristics play a role in explaining mortality clustering.

In the fourth approach, the expected number of families experiencing a certain number of deaths is compared with the observed number, to determine the extent to which excess deaths occur in the population (Table 1, approach 4). Models of excess deaths usually assume that the probability of child death is equal for all families within a group of a given parity. Furthermore, it is not easy to integrate the effect of individual characteristics of children or the mutual influence of siblings on each other's survival chances. At the same time, these models take the binomial chance distribution and parity into account most explicitly and allow for comparisons between populations exhibiting different fertility patterns.

The choice of a particular analytical model influences both the level at which we look at mortality clustering and the definition of mortality clustering and, because of model assumptions, it is also closely related to the potential explanatory characteristics and mechanisms that can be addressed. While sequence models offer insights mainly into factors at the child level (such as birth order and interactions within the sibling set, including survival of the preceding sibling), random intercept models and threshold models allow us to pinpoint factors that are mainly relevant for differences between families, as do models that are based on excess deaths. In the following sections, I discuss the four approaches as distinguished above in terms of the results on mortality clustering that have been produced. This discussion is organized according to the three different analytical levels: the child, the family, and the community. In the third (next) section I discuss research that predominantly gives insight into child-level predictors of early-life mortality; in the fourth section I move on to discuss studies providing insights into family-level predictors, including threshold and heterogeneity approaches; and in the fifth section I discuss studies that yield insights into the sources of community-level variation. An overview of all publications on mortality clustering, including a summary of their approach and main results, can be found in Table 2. In a final section I discuss new questions and directions for research, including some innovative methodological approaches to mortality clustering.

The role of scarring: influence of the death of an older sibling

While mortality clustering refers to high mortality in certain families, many current studies have focused, paradoxically, on the survival of individual children. The focus has been on the effect of the death of a sibling, usually an older sibling, on the survival of an index child (see Table 2, studies marked as approach 1). The relationship between survival of siblings has been interpreted as a scarring effect, which can be attributed to several mechanisms. First, exposure to infectious disease and death may

Table 2 Overview of publications on early-life mortality clustering

Authors and year	Population[a] and period[b]	Number of families	Number of children	Age group[c]	Approach[d]	Main findings on early-life mortality clustering
Alam (1995)	DSS, Teknaf, Bangladesh (1983–84)	NA	3,729	0–35 months	1	Birth spacing and survival are more strongly associated if the sibling survives
Alam and David (1998)	DNFS, Matlab, Bangladesh (1977–90)	NA	32,650	0–35 months	3	Adjacent siblings' mortality risk is affected by death of older sibling: neonatal mortality is higher; toddler mortality lower
Alter et al. (2001)	Sart, Belgium (1812–1900)	NA	918; 839; 469; 113	1–14; 15–29; 30–54; and 55+ years	1, 2	Correlation between siblings in mortality weakens after age 15 and disappears after age 30; some evidence for acquired immunities
Arulampalam and Bhalotra (2006)	DHS, Uttar Pradesh, West Bengal, and Kerala (–1998)	2,340–7,297	5,950–29,937	Infants	1	Mortality clustering in families is reduced by availability of contraceptives
Arulampalam and Bhalotra (2008)	DHS, 15 Indian states (–1999)	2,340–9,370	NA	Infants	1, 2	Scarring effects and evidence for sex preferences found
Beise and Voland (2002)	Krummhörn, Germany (1720–1874)	NA	3,530	Children	2	Maternal but not paternal grandmothers improve likelihood of survival. Timing of death indicates that causes are exogenous
Bhalotra and van Soest (2008)	DHS, India (–1999)	7,286	29,747	Neonates	1	Scarring effects and some proof for replacement behaviour found
Bolstad and Manda (2001)	DHS, Malawi (–1992)	2,911	4,838	Infants	2, 1	No scarring effect found if family risk is taken into account. Risk factors: early succeeding conception and short breastfeeding duration
Curtis et al. (1993)	DHS, Brazil (–1986)	2,308	4,752	Postneonates	2, 1	Birth intervals are strongly related to survival
Curtis and Steele (1996)	DHS, Bolivia, Kenya, Peru, and Tanzania (–1989)	4,754–5,739	11,586–13,134	Neonates	2, 1	Biological mechanisms are likely because of similar strength of family associations in neonatal mortality in four different populations
Das Gupta (1990)	Punjab, India (1980–90)	About 1,800	1,520	Children	1, 3	Risk factors: short birth intervals, short breastfeeding duration, and birth weight; socio-economic status and childcare abilities matter

Das Gupta (1997)	Punjab, India (–1984)	674	NA	Children	3, 4	Short birth intervals appear to be the effect of mortality. Evidence for clustering found especially in low status and low education groups
Edvinsson et al. (2005)	DDB, Sundsvall and Skellefteå, Sweden (1803–1900)	20,005	133,448	Infants	3	Risk factors: social status, number of marriages, and earlier stillbirths
Guo (1993)	Six communities in Guatemala (–1976)	851	3,120	Children	2, 1	Relatively small effect of households on mortality found beyond household economic status and mothers' education
Gyimah (2009)	DHS, Ghana (–1998 and –2003)	3,540	4,938	Children	2	Higher risk found for older children in polygynous households
Holmberg and Broström (2012)	DDB, Skellefteå, Sweden (1831–90)	8,062	37,074	Infants	4, 2	Smaller effect of scarring found once death clustering is taken into account
Hussain et al. (2001)	DHS, India (–1993) and Pakistan (–1991)	5,447; 3,993	17,531; 14,050	Children	2	Consanguineous marriages play a significant role in infant mortality
Ikamari (2000)	DHS, Kenya (–1989)	NA	16,426	Children	1	Sibling death risks are correlated, especially in infancy
Janssens et al. (2010)	Twente, The Netherlands (1875–99)	163	733	0–24 years	1	Higher mortality found for girls in late childhood than for boys
Janssens and Pelzer (2012)	HSN, two cities in The Netherlands (1900–30)	353	1,556	Children	2	Occupation of the mother is unrelated; older mothers beneficial
Kippen (2011)	Tasmania (1852–57)	NA	NA	All ages	3	Deaths within families often occur closely spaced together in epidemic years
Kippen and Walters (2012)	Sart, Belgium (1812–1900)	NA	2,123	Children	1	Evidence found for sibling competition over resources and for death clustering
Kuate-Defo and Diallo (2002)	Africa-wide DHS data (–1978/1998)	NA	NA	Children	3	Preceding child death, contraceptive use, and birth spacing affect risk
Lalou (1997)	PRDH, Quebec, Canada (1621–1730)	NA	NA	Infants	4	Wet nursing and difficult deliveries due to maternal health given as explanations
Last (1992)	Farmstead in Nigeria (–1992)	30	131	All ages	Qualitative	Replacement of departed child may lead to subsequent elevated mortality risk due to timing of next birth and behavioural factors
Lindkvist and Broström (2006)	DDB, Skellefteå, Sweden (1831–90)	59,757	133,448	Infants	2	Intergenerational transmission of mortality clustering found, especially for girls

(*Continued*)

Table 2 Continued.

Authors and year	Population[a] and period[b]	Number of families	Number of children	Age group[c]	Approach[d]	Main findings on early-life mortality clustering
Lundevaller and Edvinsson (2012)	DDB, Skellefteå, Sweden (1860–1900)	4,943	23,067	Perinates	4	Rh disease explains more than a third of perinatal mortality clustering
Lynch and Greenhouse (1994)	DDB, Sundsvall, Sweden (nineteenth century)	5,754	20,626	Children	1	Scarring effect found, especially for adjacent siblings. Increased risk of death already present for oldest children in family before death of siblings
Madise and Diamond (1995)	Malawi (–1988)	3,043	6,258	0–15 years	1	Effect found of socio-economic factors and short birth intervals, controlled for mortality clustering in the household
Manda (1998)	DHS, Malawi (–1992)	2,650	3,927	Infants	2	Biological and socio-economic factors contribute to death clustering in families; no significant community variation in infant deaths
McMurray (1997)	DHS, Burundi, Uganda, and Zimbabwe (1987–89)	2,529–3,433	11,078–16,075	Children	3	Includes several regions and finds that clustering is lower in urban and malaria-free regions, and higher for mothers living in poor conditions
Miller et al. (1992)	DNFS, Matlab, Bangladesh (1975–80); Cebu, The Philippines (1983–86)	NA	1,755; 3,029	0–24 months	3	Children born after short birth interval have higher risk of death in the first two years of life, even after accounting for gestation. This is not explained by shorter birth intervals in families with a high mortality risk for children
Myntti (1993)	Village in Yemen (1988–89)	16	NA	All ages	Qualitative	Socio-economic factors, lack of social support, and attitudes are relevant
Nault et al. (1990)	PRDH, Quebec, Canada (1608–1729)	?	17,010	Infants	4, 3	Birth intervals are related to mortality due to breastfeeding and maternal depletion
Nonyane et al. (2013)	Sylhet, Bangladesh (2001–05)	13,457	24,485	Neonates	4, 2	Random household variation is almost completely due to between-mother effects and largely explained by child and mother characteristics

Omariba et al. (2007)	DHS, Kenya (–1998)	5,716	23,348	Children	2	Biodemographic factors are more important in infancy than in childhood
Omariba et al. (2008)	DHS, Kenya (–1998)	5,717	23,351	Infants	4, 1, 2	Unobserved heterogeneity and scarring both play a role in infant mortality
Pakot (2015)	Transylvania, Romania (1850–1939)	1,883	8,841	Infants	1	Longer birth intervals are related to a lower mortality risk, especially for neonates
Pavard et al. (2005)	PRDH, Quebec, Canada (1625–1759)	NA	58,365	Postneonates–15 years	2	Mother's death negatively affects survival, especially for older girls
Pelzer and Janssens (2014)	HSN, four cities in The Netherlands (1880–1920)	656	2,682	Infants	1	Mortality is more strongly determined by regional differences than religion
Quaranta et al. (2017)	Five historical European populations	381–44,429	1,445–207,071	Infants	3, 2	Intergenerational transmission of mortality clustering occurs in five European populations
Ranjan et al. (2018)	NFHS, four Indian states (–2006)	11,992	38,392	Infants	1, 2	Effects of both scarring and mother heterogeneity found
Rao et al. (1997)	NFHS, Goa, India (–1993)	1,331	NA	Children	3	Maternal education may be important because of hygiene behaviours
Reid (2001)	Derbyshire, England (1917–22)	NA	29,537	Stillbirths and neonates	1	Stillbirths are concentrated in certain women; greater influence of external factors for neonatal mortality than for stillbirths
Reid (2002)	Derbyshire, England (1917–22)	NA	24,743	Postneonates	1	Previous child deaths affect all-cause mortality, and mortality from wasting, diarrhoea, and respiratory diseases
Ronsmans (1995)	Niakhar, Senegal (1983–89)	1,664	12,752	Children	4	Presence of known within-family heterogeneity does not explain the variance between families, suggesting an interfamily component
Saha and van Soest (2011)	HDSS, Matlab, Bangladesh (1982–2005)	25,088	64,344	Infants	1	Scarring works through birth intervals, which become shorter after the death of a preceding infant

(*Continued*)

Table 2 Continued.

Authors and year	Population[a] and period[b]	Number of families	Number of children	Age group[c]	Approach[d]	Main findings on early-life mortality clustering
Saha et al. (2014)	HDSS, Matlab, Bangladesh (1987–2005)	NA	107,367	Neonates	2	Education of the mother is protective and low socio-economic status of the father detrimental for child mortality, especially for communicable diseases
Sastry (1997)	DHS, north-east Brazil (–1986)	1,051	2,946	Children	2, 1	Community variation in mortality is more relevant than family variation
Scalone et al. (2017)	Granarolo, Italy (1900–39)	1,214	3,968	Perinates and neonates	2	Variation between sharecroppers and 'landless labourers and non-rural workers' in mortality clustering; occupation irrelevant for perinatal mortality
Sear et al. (2002)	West Kiang, Gambia (1950–74)	NA	2,294	Children	2	Living maternal grandmother reduces child mortality among toddlers
Shah and Dwivedi (2011)	Sewa, Gujarat, India	33	NA	Neonates	3 and qualitative	Family history of high mortality is mainly related to prematurity and low birth weight causes of death
Van Bodegom et al. (2012)	North-east Ghana (2002–10)	1,703	9,288	Children	2	More variance occurs at household than village level; influence of both fathers and mothers found; socio-economic status and water sources play an influential role
Van Poppel et al. (2012)	LINKS, three provinces of The Netherlands (1812–1903)	90,000	485,303	Stillbirths and children	1, 2	More death clustering found in later periods; changes seen over time in effect of stillbirth on death of index child
Vandezande (2012)	COR*, Antwerp, Belgium (1846–1905)	322–376	1,222–1,826	Stillbirths and infants	2, 3	Death risks are correlated between generations in maternal and paternal line. Age at death is relevant: late neonatal deaths are most predictive
Vandezande (2013)	COR*, Antwerp, Belgium (1846–1905)	406	1,634	Postneonates		Postneonatal deaths are transmitted across generations in the paternal line

Vandezande and Matthijs (2013)	COR*, Antwerp, Belgium (1846–1905)	376	1,826	Infants	3	Evidence found for intergenerational mechanisms; these are not due to reproductive behaviour, social class effects, or persistent regional effects
Willführ and Gagnon (2012)	Krummhörn, Germany (1720–1874) PRDH, Quebec, Canada (1670–1720)	6,445 13,932	29,935 115,013	Infants	4	Stronger concentration of infant deaths found in Quebec; possibly explained by better maternal quality in populations with low marriage rate
Zaba and David (1996)	Census of Kenya (–1979)	194,198	691,546	Infants	4	Parity and age explain part of the clustering of deaths; more variation at higher ages and parities
Zenger (1993)	Matlab, Bangladesh (1968–73)	2,832	7,304	Neonates	1	Neonatal deaths are most strongly associated for adjacent siblings; birth spacing effect on mortality is stronger if sibling survived

[a]COR*, sample from Antwerp based on the first three letters of the family name; DDB, Demographic Database; DHS, Demographic and Health Survey; DNFS, Determinants of Natural Fertility Survey; DSS, Demographic Surveillance System; HDSS, Health and Demographic Surveillance System; HSN, Historical Sample of the Netherlands; NFHS, National Family Health Survey; LINKS, Linking System for Historical Family Reconstitution; PRDH, Programme de recherché en démographie historique.
[b]Where only end of observation period is provided (e.g., –1986), observations are cross-sectional, based on retrospective survey or census data.
[c]Infants refers to the age group 0–1 and children to the age group 0–5; other age ranges are specified.
[d]For an overview of approaches, methods, and their characteristics, see Table 1. Approaches are listed in order of predominance.
Note: NA refers to 'not applicable'.

result in damage to the bodies or immune systems of surviving siblings (Bengtsson and Lindström 2000; Alter et al. 2001; Barker et al. 2002; Finch and Crimmins 2004). Second, bereaved parents may be depressed, which can have detrimental effects on the remaining children because of suboptimal caretaking. Third, child deaths are related to short birth intervals through two mechanisms: after the death of an infant, breastfeeding is interrupted, which leads to faster restored fecundity (Nault et al. 1990) and, also, parents may decide to replace the lost child quickly (Arulampalam and Bhalotra 2006, 2008; Bhalotra and van Soest 2008). In turn, short birth intervals lead to higher maternal and child mortality, as it may result in maternal depletion, which increases the risk of pregnancy complications, preterm delivery, and low birth weight (Curtis et al. 1993; Omariba et al. 2008).

Using this approach, it has been shown that neonatal deaths (those between birth and the 28th day after birth) and postneonatal deaths (those later in infancy) are the most strongly related between adjacent pairs of siblings (Zenger 1993; Curtis and Steele 1996). However, in families that will eventually lose many children, this relationship is present even early on: during the years in which the family is still small, the oldest children already have increased likelihoods of death (Lynch and Greenhouse 1994). One of the explanations for scarring effects between pairs of siblings is short birth intervals (Saha and van Soest 2011). As just discussed, closely spaced births are related to maternal and early-life child mortality. However, research has shown that short birth intervals are detrimental if the preceding sibling survives, but that these effects are weaker if the preceding sibling has died (Zenger 1993). In addition, death clustering is more pronounced among higher parity women (Zaba and David 1996). These results indicate that the effect of birth intervals may be partially explained by resource competition (Kippen and Walters 2012) or transmission of infectious disease between closely spaced siblings or in crowded households (Das Gupta 1990; Alam 1995; Kippen 2011). Finally, some authors have found that the effect of scarring appears to be dependent on the sex of the preceding sibling, which could indicate that son preferences play a role in the correlation between sibling deaths (Arulampalam and Bhalotra 2008).

Later research in this vein took not only the effect of sibling survival into account, but used sequence models that were able to control for family-level heterogeneity, thus accounting for the fact that observations within families are not independent. These studies have shown that scarring and unobserved heterogeneity both play a role in explaining early-life mortality, emphasizing that there are differences between families in mortality levels and also that the death of a sibling has a causal effect on survival of the next sibling (Arulampalam and Bhalotra 2006, 2008; Omariba et al. 2008). However, once clustering in the family is taken into account, scarring seems to play a much more modest role (Bolstad and Manda 2001; Saha and van Soest 2011; Holmberg and Broström 2012). Saha and van Soest (2011) have suggested that, when keeping heterogeneity constant (i.e., controlling for death clustering), the death of a sibling scars remaining siblings, increasing the risk of death by 29 per cent in a contemporary population in Bangladesh, mainly through a decrease in the interval to the next birth. This study is a welcome contribution among the methodologically more advanced contributions to scarring research, as it has identified a pathway by which scarring influences the family. In itself, scarring is a black box, as it does not clarify the mechanisms by which a first death causes further deaths and does not reveal why the death of a child is so detrimental to their brothers and sisters. Furthermore, it tells us little more than that a first death is related to higher odds of further deaths and should thus function as a red flag for local governments and healthcare workers, among others. Still, the causes of the initial death remain unknown and we learn little about the families in which deaths occur.

An additional problem of research into the role of scarring is that the econometric methods that are commonly used to determine the impact of scarring in the family have strict model assumptions. First, models that take both unobserved heterogeneity between families and a scarring effect into account use a Markov approach. Markov models assume that only adjacent children are influenced by the death of a sibling (Omariba et al. 2008), but that may not be entirely true. The mechanisms that cause a relationship between the mortality of adjacent brothers and sisters may influence non-adjacent children as well as adjacent children. For instance, all children are likely to suffer from the consequences of parental depression or maternal depletion, not just younger or adjacent children. In addition, it has been shown that effects of sibling death on non-adjacent siblings do exist, although the effects are largest for adjacent children (Alam 1995). Omariba et al. (2008) included random effects in their models to capture the influences of non-adjacent children; however, this solution assumes that the non-adjacent effects influence all other children in a similar fashion. Second, for the assumption of sequence to hold, older

siblings must die before younger siblings. Scarring research must focus on infant mortality, for this assumption will not always be met if research is extended to child mortality. At the same time, death among younger siblings may influence the survival of both older and younger siblings. Finally, as discussed above, death clustering implies that there is a stronger concentration of early-life deaths in families than would be expected. As the scarring models often include only the effect of death of the adjacent sibling, it is unclear to what extent there is an unusual concentration of deaths in these families. For an exception and a discussion of higher-order Markov models to solve this issue, see Omariba et al. (2008).

A caveat should be made here, however. Not only is the mechanism behind scarring effects often unclear and the methods not completely sufficient, there have also been favourable developments in infant and child mortality rates in less developed countries over recent years. In addition, families have become smaller on average (UNICEF et al. 2014). Therefore, scarring effects on siblings will have become less relevant for health and survival. This suggests that part of the decline in mortality in modern-day less developed countries is caused by a self-reinforcing mechanism: because fewer children die, the health of fewer siblings is damaged, causing further reductions in child mortality. Why variation exists between families in the total experience of early-life mortality, then, becomes an even more important question.

To conclude this section, research into the effect of the death of a sibling on individual survival has generated important insights into the antecedents of excess child loss in the past and in the developing world. At the same time, the frequently cited explanation—scarring—remains a 'black box' explanation, as the mechanisms through which brothers and sisters are moulded by the death of their sibling remain unclear in most research. Furthermore, research has neglected the opportunity to assess how the effects of sibling deaths differ between families of varying backgrounds. Existing research efforts can relatively easily be extended to do this, for instance, by the application of simulation models or by including family-level heterogeneity in models and research reports.

Characteristics of families and mortality clustering

A second group of studies have incorporated family characteristics more explicitly in the definition and measurement of mortality clustering and in the characteristics that are linked to early-life mortality. In early research, death clustering often referred to families experiencing multiple infant or child deaths. Other research has used heterogeneity between families in their children's chances of dying in their early years and determined how much variation in early-life mortality between mothers is explained by the inclusion of explanatory characteristics. This approach allows for the inclusion of variation at both the family and community levels. Finally, researchers have used excess deaths in families at the population level to explore family-level causes of excess mortality, or used simulation models to assess whether excess deaths are explained by the family characteristics that are known determinants of early-life mortality. The main results from studies using these approaches can be found in Table 2 under approaches 2–4.

Explanations for differences in mortality risk between families have been sought in fertility behaviour, socio-economic status, childcare practices, and maternal health and biological characteristics. With regard to fertility behaviour, multiple pregnancies, especially if closely spaced, may result in maternal depletion. Furthermore, in larger families, children compete for the limited resources available to the family. Zaba and David (1996) have shown that the variability between mothers in the risk of their children dying—an indication of mortality clustering—is highest among mothers at the highest parities. Several mechanisms could explain this relationship. First, mothers who experience more child deaths tend to proceed to higher parities. Second, life history theory predicts that individuals exposed to high mortality will reproduce early, have shorter birth intervals, and invest less in their children, which is essentially predicting mortality clustering among mothers exposed to high-mortality environments (Störmer and Lummaa 2014).

Other explanations have been sought in the realms of socio-economic status and childcare practices (Das Gupta 1990, 1997). In India, death clustering has been partially explained by the caste membership of mothers, which is related to poverty and access to healthcare (Ranjan et al. 2018). Lalou (1997) found that infant mortality was higher and birth intervals shorter in the higher classes, which may be explained by their practice of using wet nurses. In addition, in the lower classes inadequate feeding patterns may have led to poor health among mothers, resulting in repeated problems during delivery, repeated stillbirths, and high maternal mortality (Lalou 1997). Janssens and Pelzer (2012) found that

mothers who had worked in factories, who at the time were assumed to turn into 'bad mothers', did not do worse than other women. Das Gupta (1997) acknowledged that there was more variety in parenting behaviour among the women with relatively low levels of education in her sample than among those with higher levels of education, resulting in higher mortality among the children of women with lower education. This implies that maternal care plays an important role in child mortality, but that there is also an overlap with socio-economic status. Although there is little evidence for a consistent relationship between socio-economic status and childhood mortality for much of the nineteenth century (Janssens and Pelzer 2012), Scalone et al. (2017) found that, in nineteenth-century Italy, sharecroppers were less likely to experience mortality clustering than landless labourers and non-rural workers. This indicates that co-resident kin—common among sharecroppers—decrease the likelihood of mortality clustering. However, it is unclear to what extent these families co-resided with their kin; moreover, economic deprivation and seasonal migration may be alternative explanations for the higher mortality levels of the landless labourers and non-rural workers. Thus, future research should attempt to distinguish carefully between the role of proximate kin and the role of socio-economic conditions.

With regard to social and biological characteristics, Pavard et al. (2005) found that maternal death decreased the chance of survival of the children affected, especially girls, which suggests that the presence of kin and maternal care both play an important role in survival. Similarly, Edvinsson et al. (2005) found that both mattered in determining risk status: remarriage after the death of a spouse (interpreted as a largely social characteristic related to instability of the family) and the experience of stillbirth (considered by the authors to most likely have been caused by biological mechanisms including health status and genetic factors) were both related to the likelihood of being a high-risk mother. Willführ and Gagnon (2012) used a Lorenz curve and Gini coefficients for the expected and observed distributions of deaths over the population, and showed that there was less clustering in Krummhörn (Germany) than in Quebec (Canada). The extent to which deaths cluster differed between recomposed and nuclear families, showing that in Quebec maternal care abilities may have played a role in mortality clustering in the families of remarried widowers (Willführ and Gagnon 2012). Furthermore, consanguineous marriages, which are common in some parts of the world, including Pakistan and India, are related to excess early-life mortality (Hussain et al. 2001). Earlier research has suggested that frailty—and hence the risk of death—is caused by genetic disposition (Vaupel et al. 1979; Vaupel 1988; Yashin et al. 1995). However, in the absence of genetic information, it is difficult to attribute remaining family variance to genetic causes. Remaining familial heterogeneity may be considered the upper limit of the influence of genes shared in families.

Furthermore, it has been shown that there is an intergenerational component to death clustering (Lindkvist and Broström 2006; Vandezande and Matthijs 2013; Quaranta et al. 2017). Parents who lost many siblings in infancy are more likely to have a high-mortality family themselves. Thus, mortality clustering is transferred between generations. The causes of high mortality in the first generation—be they social, biological, or economic—appear to be transmitted to the next generation (Vandezande 2012). Alternatively, exposure to high mortality in the family of origin itself may be a cause of faster reproduction, lower investments in children, and higher early-life mortality in the second generation, as life history theory predicts. It remains unclear, however, which characteristics explain this intergenerational transmission of mortality clustering. It is interesting to note that intergenerational transmission occurs through both the paternal and maternal lines (Vandezande 2013). The influence of fathers on family-level mortality beyond socio-economic status has been shown in only one earlier paper. Van Bodegom et al. (2012) used a polygamous community in Ghana to show that variation in mortality could be found at the paternal and maternal levels separately. Mortality clustering research commonly focuses only on mother-level variation and explanations, but these papers have shown that the father plays a larger role than is commonly assumed. At the same time, earlier work has shown that there are large differences between women in the same polygynous unions (Last 1992). In comparison with monogamous unions, the offspring experience survival disadvantages in later childhood (Gyimah 2009). Furthermore, other kin may affect survival as well, as earlier research has shown that maternal grandmothers improve the survival chances of their grandchildren (Beise and Voland 2002; Sear et al. 2002).

Family-level studies often include individual-level explanations for early-life mortality and these may interact with family-level characteristics. In earlier research, it has been shown that heirs often fare better than other children (Voland 2000) and that parents may show bias towards one sex at the expense of the other (Pavard et al. 2005;

Arulampalam and Bhalotra 2008; Janssens et al. 2010). As discussed earlier, families facing economic hardship may be more likely to lose their children and infants. However, the burden may be divided unequally among the family members. For instance, during food shortages, boys' nutritional needs may be better met than those of girls. Similarly, the dependency ratio of the family results in resource dilution: resources run thin in large households, especially if the children are still young. Who suffers from this lack of resources depends on the way in which resources are divided among household members. Next to family size, other household characteristics, including household earning capacities and parental competence, are also likely to change over the life course of the family. The effects of these household characteristics on children depend on their sex and birth order, as these are related to their responsibilities within the household. Thus, household characteristics impact on the likelihood of death among infants and children in a complex fashion and, moreover, may change over time.

Interaction models focusing on both child and family characteristics may help us to achieve a more thorough understanding of these processes, by explaining family heterogeneity in mortality risks. However, models using random variation between families assume that all children within the household are influenced in a similar fashion. By incorporating interaction terms between family characteristics (including their risk status) and individual characteristics, these processes may be explored in future research.

To summarize, a focus on mother heterogeneity and familial risk status has several advantages: it clarifies why families differ in the likelihood that mortality occurs among their infants and children, and it enables competing explanations to be tested simultaneously (Edvinsson et al. 2005). Furthermore, community effects can be modelled next to family effects and researchers may assess how much random variation between families is explained by their models. At the same time, when the focus is on variation between families, it remains unclear to what extent family differences in mortality levels reflect chance fluctuations within communities. While social, economic, and biological mechanisms have all been shown to play a role in mortality clustering, it is still not clear (1) whether the same explanations apply for clustering in the past as for clustering in modern-day less developed countries; (2) which explanations matter most; and (3) which children are most at risk in high-risk families.

Mortality clustering in populations

Mortality clustering as a phenomenon shows that mortality is concentrated in families, affecting the early-life survival chances of their children, and manifests itself at the population level as an excess concentration of deaths in some families. Although the methodological angles discussed in the previous two sections focus on the child and family levels, respectively, most papers use population-level measures to show the extent to which mortality clustering manifests itself in the populations under examination. Several researchers have taken a step beyond that and managed to explore to what extent death clustering in the population can be explained by the characteristics of children and families. Results from these papers can be found in Table 2, under approach 4. To determine the relevance of the explanations given and whether we should continue to search for new explanations, we need to assess whether there is still excess clustering in families (exceeding the clustering as predicted by chance) after taking our explanations into account. In other words, we should return to the population level. Holmberg and Broström (2012) have proposed using simulation models to find the extent to which mortality clustering in the population is explained. Using this method, they first account for family size and use the binomial chance distribution to find the extent to which deaths cluster in a population. Second, models with random effects are estimated to analyse the causes of mortality clustering. Third, these explanatory variables are incorporated in simulation models to find the expected number of deaths, which is compared to the observed number of deaths. Thus, the extent to which there is still excess mortality after including known explanations for mortality clustering is assessed.

While most studies into early-life mortality clustering assume that it can be found in any population, death clustering seems to play a larger role in some communities than in others (Van Poppel et al. 2012; Vandezande 2012). For north-east Brazil, Sastry (1997) has shown that taking community-level variance in mortality into account explains almost all variance between families in the number of children that succumb, suggesting that mortality clustering may be overestimated in some contemporary research. Strikingly, Saha and van Soest (2011) used a random cluster analysis to show that in Bangladesh the main source of variation in early-life mortality is found in families, not in communities. Similar results were found by Manda (1998). This means

that, here, families play a more important role in death clustering than their communities.

The causes of community variation have rarely been the subject of research. In some rare examples addressing differences between populations, it has been proposed that some populations are culturally quite homogenous and socio-economically comparatively equal, which results in relatively little inequality in mortality between families (Curtis et al. 1993; Guo 1993). For instance, Guo (1993) found only small differences between families in Guatemala after accounting for varying income levels and suggested that, in high-mortality populations with relatively low socio-economic inequality, unexplained variance is low, because mortality resulting from poverty and community effects is still very high. In modernizing societies, on the other hand, varying levels of access to healthcare and new health interventions could increase inequalities between families, as better-off families are better equipped to take advantage of these new services (Saha and van Soest 2011). However, other studies have found unexplained mortality clustering in relatively high-mortality populations, such as nineteenth-century Netherlands (Van Poppel et al. 2012), indicating that increasing inequalities after the start of the epidemiological transition are not the sole cause of mortality clustering. Spatial factors including access to safe sources of water (Van Bodegom et al. 2012) and rural–urban differences (Vandezande 2012) have been shown to play a role in mortality clustering in high-mortality contexts. In addition, cultural factors such as attitudes towards breastfeeding may play a role (Pelzer and Janssens 2014). In future research, both population- and family-level characteristics need to be taken into account to clarify whether significant familial mortality clustering exists, or whether populations play a more important role than previously thought.

Another advantage of a stronger focus on the role of community characteristics in mortality clustering would be the insight delivered into the contextual determinants of changing mortality patterns, such as the causes of the epidemiological transition. Generally, historical demographers and international development researchers have failed to make use of the opportunity to explain the epidemiological transition better by focusing on the changing characteristics over time of families who lose their infants and children. In other words, the experiences of families who have carried the burden of mortality—or escaped early-life mortality—sharpen our understanding of mortality in the past and in the modern-day developing world, especially if we take longitudinal changes in the causes of clustering into account. Furthermore, the role of context in high familial mortality helps to explain why mortality clusters in some families but not in others. Why is it that some families are especially vulnerable to adverse circumstances and how does that change over time? To explore that, we need to research the interaction between family and community characteristics. Thus, we need to find out which families experience excess deaths and why. This strategy may help to explain why death clustering is more common in some populations than in others and to identify the determinants of death clustering at the family level.

The way forward and concluding remarks

After 25 years of research on infant and child mortality clustering, important insights into the causes of early-life mortality clustering have been generated. The research has shown that in populations around the world and over time, there are differences between families in the degree of risk of mortality among their children. It has become clear that, taking the varying risk between families into account, the death of an older child harms the survival chances of younger children in that family and that fertility behaviour, earlier stillbirths, remarriages, and socio-economic status all explain mortality clustering to some extent. However, many questions on the nature of the phenomenon remain and several promising avenues in mortality clustering research have been neglected in research to date. In that sense, death clustering remains a black box: to open it, we need to concentrate on the pivotal role played by the families in which deaths cluster.

First, social, economic, and biological mechanisms have each been shown to play a role in mortality clustering; and birth intervals, birth order, age of the mother, and other demographic characteristics also matter, as does an effect of scarring from sibling death. However, it remains unclear whether the same explanations apply for clustering in the past as for clustering in modern-day less developed countries, as systematic comparisons are lacking. In high-mortality societies, spatial differences appear to play a more important role, whereas in transitioning and low-mortality societies, inequalities between families are more important. These community-level characteristics are an underexplored but promising avenue for future research. Modernization of economies eventually benefits most individuals, but especially during the initial phases of modernization, the benefits may be limited to some families; others

may see no benefits and modernization may even have a negative influence for some, increasing inequalities in early-life mortality. The interaction between shifting conditions over time and the extent to which deaths cluster in families is an important avenue for future research, one that has remained unexplored so far. Furthermore, because the levels of child and infant mortality differ between populations, as does the extent to which deaths cluster in families, we should develop strategies to take the varying levels of mortality between populations into account in our models. Currently, part of the inequality in mortality shown by our models may be caused by the accidental attribution of community-level variance to the family level.

Second, which explanations matter most and for whom—which children, of what age—remains as yet unknown. More innovative questions may be asked if researchers focus on the interactions between the analytical levels that were addressed earlier in this paper. The child's characteristics may interact with those of their family, which may be more beneficial to some children than to others. Similarly, the community environment in which the family lives may be more challenging for some families than for others and, at the same time, have a more detrimental effect on some children within that family than on others, such as on later-born children or girls. These interactions between analytical levels have not yet been taken into account in a systematic way. Finally, it is not yet known whether the causes of mortality are similar for high- and low-mortality families. Do the same characteristics put infants and children at risk in both high- and low-risk families, or are there specific explanations for mortality in high-risk families? These questions are essential for making progress in the field of early-life mortality clustering in families.

Third, at the centre of most death clustering research, we find the nuclear family; family relationships beyond parents and children in a nuclear household are seldom considered in such research. However, it has been shown that intergenerational transfers of mortality clustering may apply if the causes of mortality clustering are transferred between generations (Vandezande 2012). Furthermore, conjugal and extended families may be more important in populations in transition than previously acknowledged in demographic and development research. This means that adults (other than biological parents) who co-reside with families should not be excluded from our theories and models. In addition to vertical (intergenerational) similarities in mortality clustering, horizontal similarity to siblings may enhance understanding of the phenomenon: siblings may share learned behaviour, socio-economic status, environments, and biological difficulties, resulting in similar patterns of mortality among their children. Finally, although most research calculates rates of mortality clustering based on mothers, it has been shown that fathers may play an important role too and, therefore, should be considered in more detail. The practical implications of these suggestions for innovative research questions concern the tailoring of policy interventions. If mortality clustering is socially contagious or inherited, policy interventions aimed at combating early-life mortality should focus on the family members of people who experience mortality clustering among their children, rather than on all families or only on the mothers and siblings of parents in whose families mortality clusters.

The limits for answering questions such as these have often been set by the availability of suitable data. In sources concerning developing world populations, such as the Demographic and Health Survey (DHS) and digitized censuses, individuals are surveyed about their reproductive histories and household characteristics. These databases are often rich in detail, with intricate insights into the socio-economic status and knowledge of parents and the conditions surrounding births. Furthermore, similar census and survey questions are asked in many countries, allowing for international comparisons. Most research on death clustering focuses on one or two regions only, so including more regions in research, especially using similar data sources and approaches, would enhance our understanding of the phenomenon (see McMurray 1997 for an example). Several drawbacks of these data should be mentioned as well. The retrospective reconstruction of reproductive histories is sensitive to cultural differences, social desirability, memory, and emotional sensitivities. In addition, often only women are interviewed, while the paternal perspective is neglected. Finally, women who did not survive their reproductive period are not included in cross-sectional surveys.

In historical demography, the use of censuses and population or parish registers has been common. These data sets often give insight not only into an individual's life course, but also into those of their parents, grandparents, and offspring. Traditionally, these data sets were often small in scope, concerning only one village or region over a short period of time. Therefore, they often did not contain information about close relatives residing outside the region of

concern, or (if a sample was taken) who did not live in the same household, therefore limiting the scope of research based on these data sources. In recent years, the digitization of sources, including civil certificates and population and parish registers, has led to the increasing availability of high-quality data sources on individuals. By linking information from various sources, the scope of historical demographic data sets has increased, allowing for comparisons between siblings, research on the linked lives of multiple generations, and comparisons of populations in multiple (sub)regions. Thus, new areas for research have been opened up by the increasing computerization of data.

Fourth, many questions remain with regard to the period of death. While some authors choose to focus on postneonatal deaths (e.g., Curtis et al. 1993; Reid 2002), others focus on neonatal deaths (see, e.g., Bhalotra and van Soest 2008; Nonyane et al. 2013). Yet others include all perinatal deaths, including stillbirths (Reid 2001; Lundevaller and Edvinsson 2012). It has been argued that the timing of death is related to its cause. Postneonatal death may be related more to factors such as kin support (Sear and Mace 2008; Sear and Coall 2011), childcare (Edvinsson et al. 2005), and other exogenous characteristics (Lalou 1997; Beise and Voland 2002), while perinatal death is assumed to be related more to factors experienced during pregnancy and to health and genetic problems (Lalou 1997; Reid 2001; Vandezande 2012; Pakot 2015). For instance, Reid (2001, 2002) found that a woman's history of child deaths was related to further postneonatal mortality of her infants, especially that caused by wasting, diarrhoea, and respiratory diseases. At the same time, her history of stillbirths and miscarriages affected the likelihood of repeated stillbirths and miscarriages in the future. This implies that early-life mortality clustering in some women is related to endogenous or perinatal factors, whereas in other women it may be related to exogenous factors including behaviour.

However, this argument does not always hold: social and cultural practices may play a large role in the first weeks of life as well (Nonyane et al. 2013). For instance, in nineteenth-century Iceland, the incidence of neonatal tetanus was extremely high through infection of the umbilical stump (Garðarsdóttir 2002). Similarly, Lalou (1997) showed that neonatal mortality peaked during an epidemic of smallpox. In the literature, a systematic approach to the timing of death is often lacking and infant deaths seem to be more commonly addressed in the field than early-life child mortality (see Table 2). A way forward may be found following research that addresses the influence of sibling mortality in several periods of early life separately. For instance, Alam and David (1998) have shown that infant deaths decrease the likelihood of toddler deaths in families, but toddler deaths increase the likelihood of infant deaths, illustrating that the ages at death of siblings are related. Alter et al. (2001) took mortality over the life course into account, showing that deaths of siblings were most strongly related in early childhood, weakened after age 15, and disappeared after age 30. Children from high-mortality households tended to survive longer, pointing to an acquired immunity effect. Furthermore, research into the causes and timing of death of children in high-mortality households may help us to disentangle whether causes are behavioural (such as gastrointestinal and external causes of death) or biological (such as genetic and cardiovascular diseases), or should be understood as caused by the transmission of disease between siblings, for instance, if multiple siblings succumb to the same cause of death in a short time frame (Kippen 2011). In addition, researchers could investigate the extent to which the factors related to mortality clustering are related to specific causes of death instead of all-cause mortality. Causes of death have only rarely been incorporated in research into the mortality of siblings (for examples, see Reid 2001, 2002; Kippen 2011; Saha et al. 2014).

Finally, early-life mortality clustering is an exciting field of research, with many avenues yet to be explored. Addressing the questions explored in this review of the literature on mortality clustering would provide us with important insights into the demographic transition and help us to understand current and future developments with regard to early-life mortality and health in the developing world. Although data sources and models remain highly complex and challenging for researchers, asking sophisticated new questions and applying more thorough analysis to existing data should help the determinants of early-life mortality and the role of families become much clearer in the near future. The existing models and innovative research approaches that were discussed in this review can be used to help us answer these questions.

Notes and acknowledgements

1 Please direct all correspondence to Ingrid K. van Dijk, Radboud University Nijmegen, History, PO Box 9103, 6500 HD Nijmegen, The Netherlands; or by E-mail: ik.vandijk@let.ru.nl

2 This work is part of the research programme 'Genes, Germs and Resources', supported by the Netherlands Organisation for Scientific Research (NWO) [grant number 360–53–180].

References

Alam, N. 1995. Birth spacing and infant and early childhood mortality in a high fertility area of Bangladesh: age-dependent and interactive effects, *Journal of Biosocial Science* 27(4): 393–404.

Alam, N. and P. H. David. 1998. Infant and child mortality in Bangladesh: age-specific effects of previous child's death, *Journal of Biosocial Science* 30(3): 333–348.

Alter, G., M. Oris, and G. Broström. 2001. The family and mortality: a case study from rural Belgium, *Annales de démographie historique* 101(1): 11–31.

Arulampalam, W. and S. Bhalotra. 2006. Sibling death clustering in India: state dependence versus unobserved heterogeneity, *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 169(4): 829–848.

Arulampalam, W. and S. Bhalotra. 2008. The linked survival prospects of siblings: evidence for the Indian states, *Population Studies* 62(2): 171–190.

Barker, D. J. P., J. G. Eriksson, T. Forsén, and C. Osmond. 2002. Fetal origins of adult disease: strength of effects and biological basis, *International Journal of Epidemiology* 31(6): 1235–1239.

Beise, J. and E. Voland. 2002. A multilevel event history analysis of the effects of grandmothers on child mortality in a historical German population (Krummhörn, Ostfriesland, 1720–1874), *Demographic Research* 7 (13): 469–498.

Bengtsson, T. and M. Lindström. 2000. Childhood misery and disease in later life: the effects on mortality in old age of hazards experienced in early life, southern Sweden, 1760–1894, *Population Studies* 54(3): 263–277.

Bhalotra, S. and A. van Soest. 2008. Birth-spacing, fertility and neonatal mortality in India: dynamics, frailty and fecundity, *Journal of Econometrics* 143(2): 274–290.

Bolstad, W. M. and S. O. Manda. 2001. Investigating child mortality in Malawi using family and community random effects, *Journal of the American Statistical Association* 96(453): 12–19.

Curtis, S. L., I. Diamond, and J. W. McDonald. 1993. Birth interval and family effects on postneonatal mortality in Brazil, *Demography* 30(1): 33–43.

Curtis, S. L. and F. Steele. 1996. Variations in familial neonatal mortality risks in four countries, *Journal of Biosocial Science* 28(2): 141–159.

Das Gupta, M. 1990. Death clustering, mothers' education and the determinants of child mortality in rural Punjab, India, *Population Studies* 44(3): 489–505.

Das Gupta, M. 1997. Socio-economic status and clustering of child deaths in rural Punjab, *Population Studies* 51(2): 191–202.

Edvinsson, S., A. Brändström, J. Rogers, and G. Broström. 2005. High-risk families: The unequal distribution of infant mortality in nineteenth-century Sweden, *Population Studies* 59(3): 321–337.

Finch, C. E. and E. M. Crimmins. 2004. Inflammatory exposure and historical changes in human life-spans, *Science* 305(5691): 1736–1739.

Garðarsdóttir, Ó. 2002. *Saving the Child: Regional, Cultural and Social Aspects of the Infant Mortality Decline in Iceland, 1770–1920*. Reykjavik: Gutenberg.

Guo, G. 1993. Use of sibling data to estimate family mortality effects in Guatemala, *Demography* 30(1): 15–32.

Gyimah, S. O. 2009. Polygynous marital structure and child survivorship in sub-Saharan Africa: some empirical evidence from Ghana, *Social Science & Medicine* 68(2): 334–342.

Holmberg, H. and G. Broström. 2012. On statistical methods for clustering: a case study on infant mortality, *Northern Sweden, 1831–1890, Biodemography and Social Biology* 58(2): 173–184.

Hussain, R., A. H. Bittles, and S. Sullivan. 2001. Consanguinity and early mortality in the Muslim populations of India and Pakistan, *American Journal of Human Biology* 13(6): 777–787.

Ikamari, L. 2000. Sibling mortality correlation in Kenya, *Journal of Biosocial Science* 32(2): 265–278.

Janssens, A., M. Messelink, and A. Need. 2010. Faulty genes or faulty parents? Gender, family and survival in early and late childhood in the Netherlands, 1860–1900, *The History of the Family* 15(1): 91–108.

Janssens, A. and B. Pelzer. 2012. Did factory girls make bad mothers? Women's labor market experience, motherhood, and children's mortality risks in the past, *Biodemography and Social Biology* 58(2): 133–148.

Kippen, R. 2011. 'A pestilence stalks abroad': familial clustering of deaths during the Tasmanian scarlet fever, measles and influenza epidemics of 1852–1854, *Genus* 67(2): 57–75.

Kippen, R. and S. Walters. 2012. Is sibling rivalry fatal? Siblings and mortality clustering, *Journal of Interdisciplinary History* 42(4): 571–591.

Kuate-Defo, B. and K. Diallo. 2002. Geography of child mortality clustering within African families, *Health & Place* 8(2): 93–117.

Lalou, R. 1997. Endogenous mortality in New France: at the crossroads of natural and social selection, in A. Bideau, B. Desjardins, and H. Pérez Brignoli (eds), *Infant and Child Mortality in the Past*. Oxford: Clarendon Press, 203–215.

Last, M. 1992. The importance of extremes: the social implications of intra-household variation in child mortality, *Social Science & Medicine* 35(6): 799–810.

Lindkvist, M. and G. Broström. 2006. Clustered infant mortality in an intergenerational perspective, Skellefteå 1831—1890. Paper presented at the Annual Meeting of the Social Science History Association, Minneapolis, November 2006.

Lundevaller, E. H. and S. Edvinsson. 2012. The effect of Rh-negative disease on perinatal mortality: some evidence from the Skellefteå region, Sweden, 1860–1900, *Biodemography and Social Biology* 58(2): 116–132.

Lynch, K. A. and J. B. Greenhouse. 1994. Risk factors for infant mortality in nineteenth-century Sweden, *Population Studies* 48(1): 117–133.

Madise, N. J. and I. Diamond. 1995. Determinants of infant mortality in Malawi: an analysis to control for death clustering within families, *Journal of Biosocial Science* 27(01): 95–106.

Manda, S. O. M. 1998. Unobserved family and community effects on infant mortality in Malawi, *Genus* 54(1/2): 143–164.

McMurray, C. 1997. Measuring excess risk of child mortality: an exploration of DHS I for Burundi, Uganda and Zimbabwe, *Journal of Biosocial Science* 29(1): 73–91.

Miller, J. E., J. Trussell, A. R. Pebley, and B. Vaughan. 1992. Birth spacing and child mortality in Bangladesh and the Philippines, *Demography* 29(2): 305–318.

Mosley, W. H. and L. C. Chen. 1984. An analytical framework for the study of child survival in developing countries, *Population and Development Review* 10(s1): 25–45.

Myntti, C. 1993. Social determinants of child health in Yemen, *Social Science & Medicine* 37(2): 233–240.

Nault, F., B. Desjardins, and J. Légaré. 1990. Effects of reproductive behaviour on infant mortality of French-Canadians during the seventeenth and eighteenth centuries, *Population Studies* 44(2): 273–285.

Nonyane, B. A. S., E. K. Williams, C. Blauvelt, M. R. Shah, G. L. Darmstadt, L. Moulton, and A. H. Baqui. 2013. Clustering of neonatal deaths in Bangladesh: results from the Projahnmo studies, *Paediatric and Perinatal Epidemiology* 27(2): 165–171.

Omariba, D. W. R., R. Beaujot, and F. Rajulton. 2007. Determinants of infant and child mortality in Kenya: an analysis controlling for frailty effects, *Population Research and Policy Review* 26(3): 299–321.

Omariba, D. W. R., F. Rajulton, and R. Beaujot. 2008. Correlated mortality of siblings in Kenya: the role of state dependence, *Demographic Research* 18(11): 311–336.

Pakot, L. 2015. Maternal health and infant mortality in rural Transylvania. A case study of Vlăhiţa and Căpâlniţa, 1850–1939, *Romanian Journal of Population Studies* 9(1): 5–24.

Pavard, S., A. Gagnon, B. Desjardins, and E. Heyer. 2005. Mother's death and child survival: the case of early Quebec, *Journal of Biosocial Science* 37(2): 209–227.

Pelzer, B. and A. Janssens. 2014. Lovely little angels in heaven? The influence of religiously determined cultural life scripts on infant survival in the Netherlands, 1880–1920, *Historical Social Research* 39(1): 19–47.

Quaranta, L., G. Broström, I. K. van Dijk, R. Donrovich, S. Edvinsson, E. Engberg, K. Mandemakers, K. Matthijs, P. Puschmann, and H. Sommerseth. 2017. Intergenerational transfers of infant mortality in historical contexts: a comparative study of five European populations. Paper presented at the Annual Meeting of the Population Association of America, Chicago, April 2017.

Ranjan, M., L. K. Dwivedi, and R. Mishra. 2018. Caste differentials in death clustering in central and eastern Indian states, *Journal of Biosocial Science* 50: 254–274.

Rao, S. R., A. Pandey, and K. I. Shajy. 1997. Child mortality in Goa: a cross-sectional analysis, *Social Biology* 44(1–2): 101–110.

Reid, A. 2001. Neonatal mortality and stillbirths in early twentieth century Derbyshire, England, *Population Studies* 55(3): 213–232.

Reid, A. 2002. Infant feeding and post-neonatal mortality in Derbyshire, England, in the early twentieth century, *Population Studies* 56(2): 151–166.

Ronsmans, C. 1995. Patterns of clustering of child mortality in a rural area of Senegal, *Population Studies* 49(3): 443–461.

Saha, U. R. and A. van Soest. 2011. Infant death clustering in families: magnitude, causes, and the influence of better health services, Bangladesh 1982–2005, *Population Studies* 65(3): 273–287.

Saha, U. R., A. van Soest, and G. E. Bijwaard. 2014. Cause-specific neonatal deaths in rural Bangladesh, 1987–2005: levels, trends, and determinants, *Population Studies* 68(3): 247–263.

Sastry, N. 1997. Family-level clustering of childhood mortality risk in Northeast Brazil, *Population Studies* 51 (3): 245–261.

Scalone, F., P. Agati, A. Angeli, and A. Donno. 2017. Exploring unobserved heterogeneity in perinatal and neonatal mortality risks: the case of an Italian share-cropping community, 1900–39, *Population Studies* 71 (1): 23–41.

Sear, R. and D. Coall. 2011. How much does family matter? Cooperative breeding and the demographic transition, *Population and Development Review* 37(s1): 81–112.

Sear, R. and R. Mace. 2008. Who keeps children alive? A review of the effects of kin on child survival, *Evolution and Human Behavior* 29(1): 1–18.

Sear, R., F. Steele, I. A. McGregor, and R. Mace. 2002. The effects of kin on child mortality in rural Gambia, *Demography* 39(1): 43–63.

Shah, B. D. and L. K. Dwivedi. 2011. Causes of neonatal deaths among tribal women in Gujarat, India, *Population Research and Policy Review* 30(4): 517–536.

Störmer, C. and V. Lummaa. 2014. Increased mortality exposure within the family rather than individual mortality experiences triggers faster life-history strategies in historic human populations, *PloS ONE* 9(1): e83633.

UNICEF, WHO, The World Bank, and the United Nations Population Division. 2014. *Levels and Trends in Child Mortality 2014*. New York: UNICEF.

Van Bodegom, D., U. K. Eriksson, J. J. Houwing-Duistermaat, and R. G. Westendorp. 2012. Clustering of child mortality in a contemporary polygamous population in Africa, *Biodemography and Social Biology* 58 (2): 162–172.

Van Poppel, F., G. E. Bijwaard, P. Ekamper, and K. Mandemakers. 2012. Historical trends in the correlation of sibling deaths in infancy in The Netherlands, *Biodemography and Social Biology* 58(2): 87–115.

Vandezande, M. 2012. *Born to Die. Death Clustering and the Intergenerational Transmission of Infant Mortality, the Antwerp District, 1846–1905*. Leuven: Centrum voor Sociologisch Onderzoek.

Vandezande, M. 2013. Using death clustering in an attempt to reassess the impact of men on infant mortality, *WOG Working paper* (2).

Vandezande, M. and K. Matthijs. 2013. Inherited dimensions of infant mortality. Detecting signs of disproportionate mortality risks in successive generations, *The History of the Family* 18(2): 169–186.

Vaupel, J. W. 1988. Inherited frailty and longevity, *Demography* 25(2): 277–287.

Vaupel, J. W., K. G. Manton, and E. Stallard. 1979. The impact of heterogeneity in individual frailty on the dynamics of mortality, *Demography* 16(3): 439–454.

Voland, E. 2000. Contributions of family reconstitution studies to evolutionary reproductive ecology, *Evolutionary Anthropology: Issues, News, and Reviews* 9(3): 134–146.

Willführ, K. P. and A. Gagnon. 2012. Are stepmothers evil or simply unskilled? Infant death clustering in recomposed families, *Biodemography and Social Biology* 58 (2): 149–161.

Yashin, A. I., J. W. Vaupel, and I. A. Iachine. 1995. Correlated individual frailty: an advantageous approach to survival analysis of bivariate data, *Mathematical Population Studies* 5(2): 145–159.

Zaba, B. and P. H. David. 1996. Fertility and the distribution of child mortality risk among women: an illustrative analysis, *Population Studies* 50(2): 263–278.

Zenger, E. 1993. Siblings' neonatal mortality risks and birth spacing in Bangladesh, *Demography* 30(3): 477–488.