Error and Bias in Under-5 Mortality Estimates Derived from Birth Histories with Small Sample Sizes

Laura Dwyer-Lindgren[*], Emmanuela Gakidou, Abraham Flaxman, and Haidong Wang

Institute for Health Metrics and Evaluation, University of Washington

Abstract

Complete and summary birth histories collected in surveys and censuses are routinely used for estimating under-5 mortality at the national level. Subnational and stratified analyses are also desirable, but the usefulness of estimates of under-5 mortality derived from small samples is unknown. We perform a data-based simulation study using Demographic and Health Survey data to quantify the expected magnitude and direction of error associated with estimates derived from birth history data with small samples. We find all methods are prone to high levels of error at small sample sizes but performance is not uniform across methods. We also perform stratified analyses to test how the performance of each method varies by level of true mortality and time prior to survey. We find that for summary birth history methods and complete birth history methods that involve considerable smoothing over time performance varies by level of true mortality and time prior to survey.

Background

Under-5 mortality, the probability of death before age 5 (denoted ${}_5q_0$), is an important overall indicator of child health. In countries without functioning systems to continuously register births and deaths, estimates of under-5 mortality are generally derived from survey and/or census data, particularly in the form of birth histories where women are asked for information about the survival of their children.

Birth history data are routinely used for estimating mortality at the national level. It is often of interest, however, to estimate under-5 mortality at a subnational level or to stratify by some other characteristic (e.g. income or maternal education). Such subnational or stratified analyses are complicated by small sample sizes: in the case of surveys in particular, the sample size for a given subnational unit or stratum is often quite small and it is not apparent if the estimates derived from these limited data are useful.

Two different types of birth histories are routinely collected. In a complete birth history, women are asked for information about the date of birth and, if applicable, the age at death of each child they have given birth to. Because complete birth histories contain information about dates and ages for individual children they allow for direct calculation of under-5 mortality. In a summary birth history, women are asked only about the total number of children they have given birth to and the number of these children who are still alive. Summary birth histories lack information about dates and ages for individual children and demographic models must be employed to estimate under-5 mortality from these data. Although complete birth histories are more straightforward to analyze they are less frequently undertaken than summary birth histories which are far less labor-intensive and time-consuming to collect.

[*]E-mail: ladwyer@uw.edu

In this paper, we aimed to determine how much error and/or bias can be expected in under-5 mortality estimates derived from both types of birth histories at various small sample sizes. To this end, we carried out a data-based simulation study using Demographic and Health Survey (DHS)[1] data wherein we treated each survey as a population with known mortality and sampled from this population to mimic surveys with small sample sizes. We examined how estimates derived from summary birth history data and complete birth history data (analyzed using several alternative methods) compared in terms of error and bias at increasingly small sample sizes. Further, we performed stratified analyses to explore in more detail how the performance of each method relates to the underlying true level of mortality and the time prior to data collection.

Methods

Data

This analysis made use of all Demographic and Health Surveys (DHS)[1] publicly available as of May 2012 which contain birth histories for all women, regardless of marital status; a total of 152 surveys in 62 countries. Table 1 provides a full listing of all DHS included in this analysis.

Birth history methods

Summary birth history method

We analyzed summary birth history data using updated models and methods described in Rajaratnam, et al.[2,3]. The combined version of the maternal age cohort, maternal age period, time since first birth cohort, and time since first birth period methods was used to generate annual estimates for the 25 years proceeding each survey.

Standard complete birth history method

To analyze complete birth history data the record for each child was expanded such that there was a record of each month that a child lived and was observed under age 5 years: this will be less than the full 60 months if the child died before age 5 or if the mother was surveyed before the child reached age 5. For each child-month of life we indicated whether the child was alive or dead at the end of the month and then assigned the child-month to the appropriate time period and age group. Time periods were non-overlapping and equally sized and were assigned starting at the time of the most recent survey and moving back in time. The ages considered were 0 months, 1-11 months, 12-23 months, 24-35 months, 36-47 months, and 48-59 months; these age groupings were designed such that mortality is expected to be reasonably constant within the age group. From these data we calculated the monthly probability of survival in each time period for each age group by calculating the proportion of child months in a given time period and age group that end with the child alive. These monthly probabilities of survival were converted to the probability of surviving the entire age interval under consideration by raising them to a power equal to the number of months in the age interval. Under-5 mortality was then calculated by subtracting from one the product of all of the age-specific survival probabilities. This process generated a single estimate of under-5 mortality for each time period which was then assigned to the midpoint of the period. Different length periods can be used with longer periods providing more pooling of information across time but also producing less frequent estimates. For this analysis, we tested periods of length 1, 2, and 5 years.

Moving window complete birth history method

As an alternative to the above, the same procedures were carried out except that instead of having non-overlapping time periods and generating one estimate per period, an estimate was generated for each year incorporating all data from a window around that year. This 'moving window' variant used each observed child-month multiple times and allowed for pooling of information across time while still producing annual estimates. Two different versions of this variant were used. In the first, all data that are within the window were treated equally, while in the second, data were weighted using triangle weights such that child-months that were closer to the year which was being predicted for were weighted more heavily than those that were further away. Different length windows can be used, with wider windows providing more pooling of information across time. For this analysis, we tested window lengths of 5 and 10 years for both variants and 20 years for the triangle-weighted variant.

Validation methods

For each survey, 'true' under-5 mortality was calculated for the total sample by applying the standard complete birth history method described above with 2-year periods. Five-hundred samples each of sizes 10, 50, 100, 500, and 1000 women were drawn without replacement from each survey for a total of 2500 samples from each survey. Estimates of under-5 mortality were derived for each of the resulting 2500 samples using the summary birth history method and each of the complete birth history methods described above. The estimates (${}_5\hat{q}_0$) for each of the 2500 samples from each method are matched to the true under-5 mortality (${}_5q_0$) by survey and year and then the error, relative error, absolute error, and absolute relative error were calculated as shown below for each sample, method, survey, and year.

$$\text{Error} = {}_5\hat{q}_0 - {}_5q_0$$

$$\text{Relative Error} = \frac{{}_5\hat{q}_0 - {}_5q_0}{{}_5q_0}$$

$$\text{Absolute Error} = |{}_5\hat{q}_0 - {}_5q_0|$$

$$\text{Absolute Relative Error} = \left|\frac{{}_5\hat{q}_0 - {}_5q_0}{{}_5q_0}\right|$$

The mean of each of these metrics was then calculated for every sample size and method across all samples and countries. The mean error and mean relative error were intended to indicate whether or not estimates from a given method are biased: since over- and under-estimates cancel in these metrics, if methods are unbiased–that is, if over-estimates and under-estimates of the same magnitude are equally likely–the mean error and the mean relative error should be approximately zero. The mean absolute error and mean absolute relative error were intended to capture the extent to which estimates of under-5 mortality can differ from true under-5 mortality; these metrics measure the magnitude of the error, regardless of the direction.

In addition to this overall analysis, stratified analyses were also carried out. First, country-years were stratified by level of true mortality–<50, 50-100, 100-150, 150-200, >200 deaths per 1,000 births–and the mean of each of the above error metrics was calculated for each method and sample size for each set of country-years. Second, country-years were stratified by the time prior to the survey–0-1, 2-3, 4-5, ..., and 24-25 years prior to the survey–and the mean of each of the above error metrics was calculated for each method and sample size for each set of country-years. These stratified analyses were meant to test if the methods perform consistently well at different levels of mortality and for different lengths of time prior to a survey.

Results

Overall performance

Figures 1-4 show the mean error, mean relative error, mean absolute error, and mean absolute relative error, respectively, observed for each method at each sample size.

The results for mean error and mean relative error suggest that all methods are close to unbiased at sample sizes of 500 or more. This is also true at smaller sample sizes for the summary birth history method. For the complete birth history methods, however, the mean error and mean relative error become noticeably negative at smaller sample sizes. This downward bias is greatest when there is less pooling, that is when period or window lengths are shorter or when triangle weights are used in place of no weights.

The mean absolute error and mean absolute relative error both increase with decreasing sample size for all methods. The magnitude of the error is substantial for all methods at sample sizes of 100 or less: no method has a mean absolute relative error less than 73%, 40%, and 29% at samples of 10, 50, and 100 women, respecitvely. In general, the summary birth history method and the complete birth history methods with long windows (10 or 20 years) perform best and the difference in performance between models increases as the sample size becomes smaller.

Stratified by true mortality

Figures 5-8 show the mean error, mean relative error, mean absolute error, and mean absolute relative error, respectively, observed for each method at each sample size stratified by true mortality level.

Most models show some tendency to under-estimate when true mortality is high and to over-estimate when true mortality is low. This is most obvious for the summary birth history method and complete birth history methods with long windows. For compelte birth history methods with shorter windows, this effect is only noticeable at sample sizes smaller than 500.

For all models the mean absolute error and mean absolute relative error differ across true mortality levels: the mean absolute error increases as mortality increases while the opposite is observed for the mean absolute relative error. The difference in the magnitude of error between the lowest mortality setting and the highest mortality setting is larger for complete birth history methods with shorter windows than for other complete birth history methods or for summary birth history methods.

Stratified by time prior to survey

Figures 9-12 show the mean error, mean relative error, mean absolute error, and mean absolute relative error, respectively, observed for each method at each sample size stratified by time prior to survey.

The mean error and mean relative error differ by time prior to survey for the summary birth history method, but there is no clear ordering in terms of time periods. In contrast, for complete birth history methods with long windows, a clear pattern emerges where the mean error and mean relative error are increasingly positive for time periods closer to the time of the survey and increasingly negative for time periods more distant from the time of the survey. For complete birth history methods with shorter periods or windows there is also a clear ordering in the mean error and mean relative error by time prior to survey, but only at the smallest sample sizes.

For all methods the mean absolute error and mean absolute relative error are larger for earlier time periods compared to time periods closer to the survey. The difference in mean absolute error and mean absolute relative error between the different time periods for a given sample size is generally the largest for complete birth history methods with short windows or periods and smaller for the summary birth history method and complete birth history methods with long windows or periods.

Discussion

All methods of analyzing birth history data considered here perform poorly at sample sizes of 100 women or less, both in terms of the magnitude of error, but also, for some methods, in terms of bias. Even at sample sizes of 500 or 1000 women, the mangitude of error for many methods is still quite high. Unfortunately the results of this analysis do not clearly point to a single best method for analyzing birth history data when sample sizes are small. The summary birth history method and complete birth history methods with long windows (10 or 20 years) perform the best overall, with less bias and considerably smaller errors than other methods. In the stratified analyses, however, we find that these same methods tend to be differentially biased at different levels of mortality and different times prior to survey. It is possible that the summary birth history method performs well overall because the models that underly this method are able to partially constrain estimates to a reasonable range. This may also be the reason why this method tends to over-predict when mortality is low and under-predict when mortality is high: it is possible that the models are biasing unusually high or low estimates towards the mean. Similarly, it is likely that the complete birth histories with long windows are among the best performers because they are able to incorporate a high degree of temporal smoothing. This temporal smoothing may also explain, however, why these methods tend to over-predict in low mortality settings and closer to the time of survey and under-predict in high mortality settings and in earlier time periods. The complete birth history methods which employ shorter time periods or windows (and consequently less temporal smoothing) also perform more consistently across different levels of true mortality and time prior to survey, at least in terms of mean error and mean relative error. Unfortunately, these methods perform the most poorly overall.

The results of this analysis suggest that birth histories in all but the largest of surveys are of limited utility for making subnational estimates or estimates for many strata. Subnational and stratified estimates are nonetheless very useful and further research into methods for using existing data sources is called for. In particular, small area models, which explicitly address small sample sizes and borrow strength both spatially and temporally, may prove useful when used with birth history data. Investment in alternative data sources, such as larger surveys or surveillance systems may also be warranted.

Table 1: Demographic and health surveys included in this analysis

Country	Survey Year(s)
Albania	2008-09
Armenia	2000; 2005; 2010
Azerbaijan	2006
Benin	1996; 2001; 2006
Bolivia	1989; 1993-94; 1998; 2003-04; 2008
Botswana	1988
Brazil	1986; 1996
Burkina Faso	1992-93; 1998-99; 2003
Burundi	1987
Cambodia	2000; 2005-06; 2010-11
Cameroon	1991; 1998; 2004
Central African Republic	1994-95
Chad	1996-97; 2004
Colombia	1986; 1990; 1995; 2000; 2004-05
Comoros	1996
Congo	2005
Congo, the Democratic Republic of the	2007
Côte d'Ivoire	1994; 1998-99
Dominican Republic	1986; 1991; 1996; 2002; 2007
Ecuador	1987
Eritrea	1995-96; 2002
Ethiopia	2000; 2005; 2010-11
Gabon	2000-01
Ghana	1988; 1993-94; 1998-99; 2003; 2008
Guatemala	1987; 1995; 1998-99
Guinea	1999; 2005
Guyana	2009
Haiti	1994-95; 2000; 2005-06
Honduras	2005-06
Kazakhstan	1995; 1999
Kenya	1988-89; 1993; 1998; 2003; 2008-09
Kyrgyzstan	1997
Lesotho	2004-05; 2009-10
Liberia	1986; 2006-07
Madagascar	1992; 1997; 2003-04; 2008-09
Malawi	1992; 2000; 2004-05; 2010
Mali	1987; 1995-96; 2001; 2006
Mauritania	2000-01
Moldova	2005
Mozambique	1997; 2003
Namibia	1992; 2000; 2006-07
Nicaragua	1997-98; 2001
Niger	1992; 1998; 2006
Nigeria	1990; 2003; 2008
Paraguay	1990
Peru	1986; 1991-92; 1996; 2000; 2004-08
Philippines	1998; 2003; 2008
Rwanda	1992; 2000; 2005; 2007-08; 2010-11
Sao Tome and Principe	2008-09
Senegal	1986; 1992-93; 1997; 2005; 2010-11
Sierra Leone	2008

South Africa	1998
Swaziland	2006-07
Tanzania, United Republic of	1991-92; 1996; 1999; 2004-05; 2009-10
Timor-Leste	2009-10
Togo	1988; 1998
Trinidad and Tobago	1987
Uganda	1988-89; 1995; 2000-01; 2006
Ukraine	2007
Uzbekistan	1996
Zambia	1992; 1996-97; 2001-02; 2007
Zimbabwe	1988-89; 1994; 1999; 2005-06; 2010-11

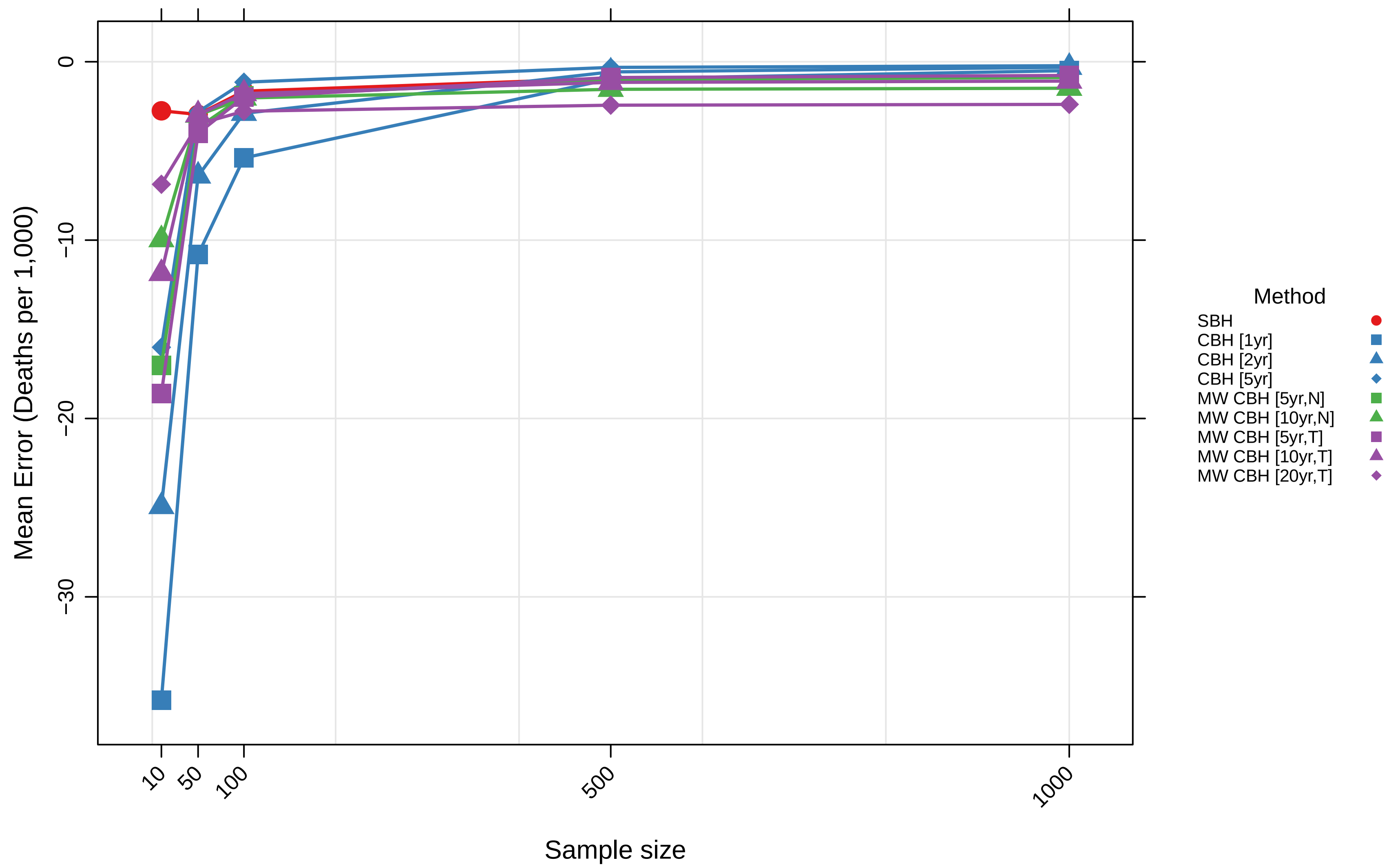

Figure 1: Mean error for all methods in the non-stratified analysis. Color of marker indicates the birth history method used, shape distinguishes different variants of each method; SBH=summary birth history; CBH=complete birth history; MW=moving window; N=no weights; T=triangle weights; number in brackets gives period or window length.

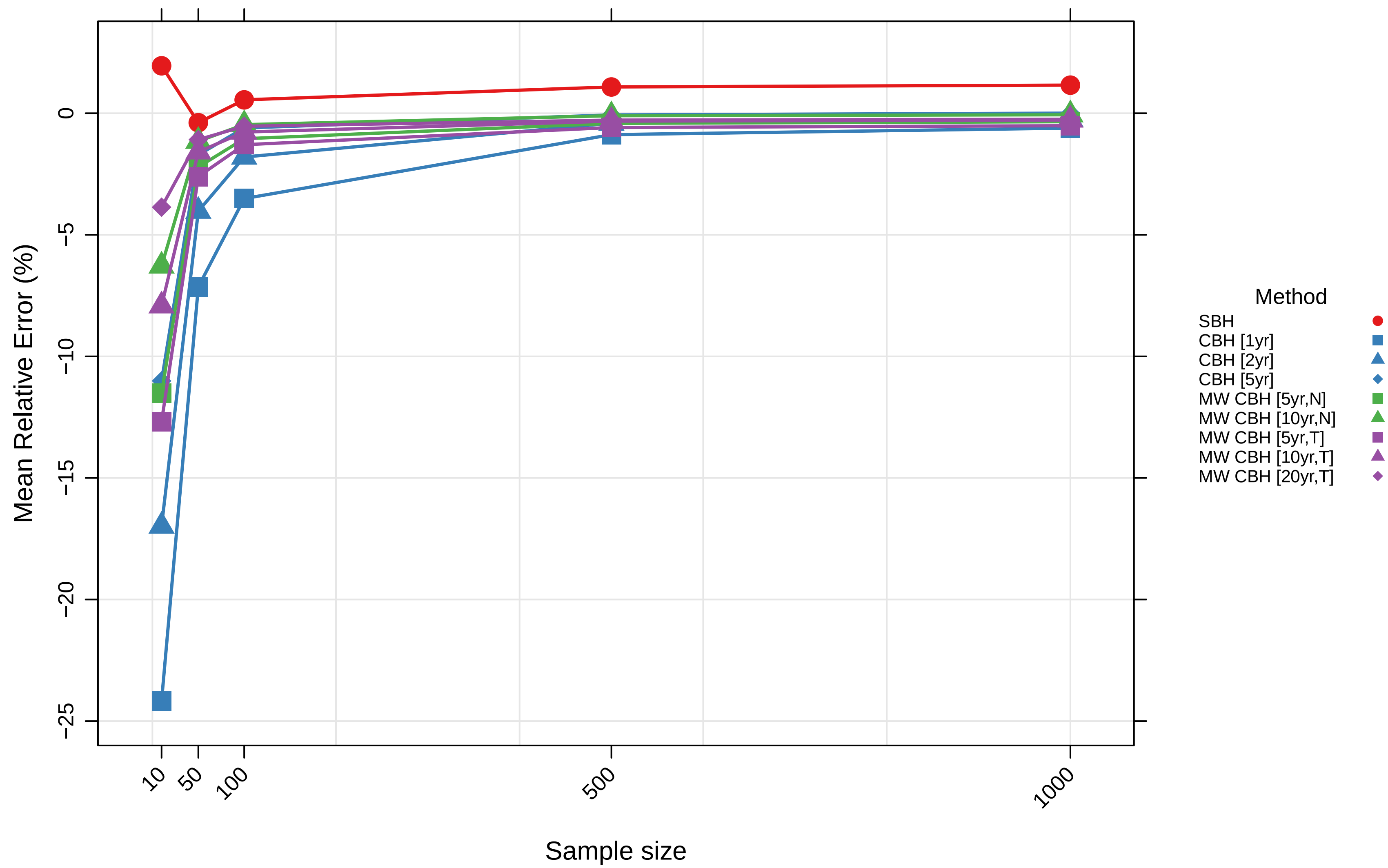

Figure 2: Mean relative error for all methods in the non-stratified analysis. Color of marker indicates the birth history method used, shape distinguishes different variants of each method; SBH=summary birth history; CBH=complete birth history; MW=moving window; N=no weights; T=triangle weights; number in brackets gives period or window length.

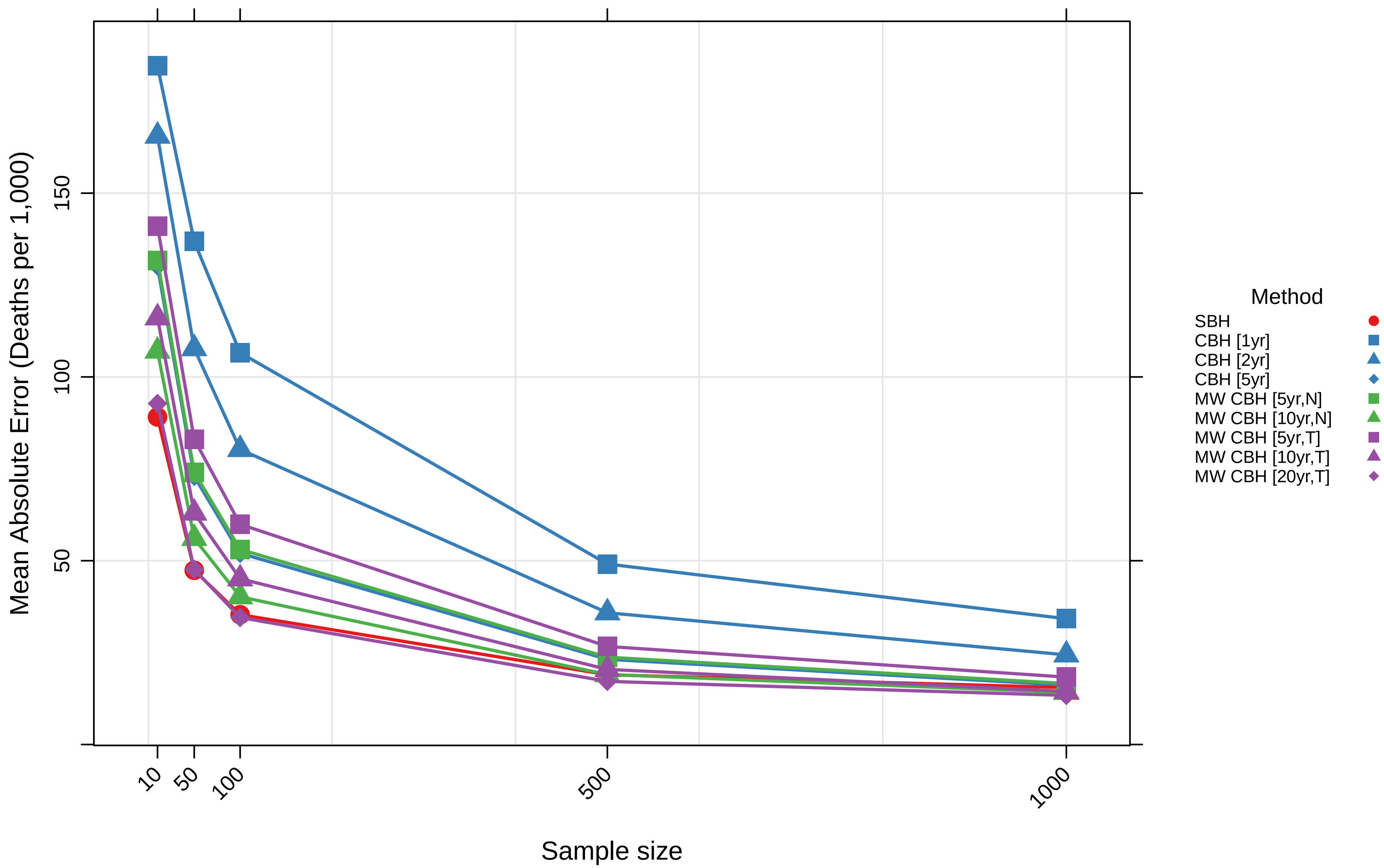

Figure 3: Mean absolute error for all methods in the non-stratified analysis. Color of marker indicates the birth history method used, shape distinguishes different variants of each method; SBH=summary birth history; CBH=complete birth history; MW=moving window; N=no weights; T=triangle weights; number in brackets gives period or window length.

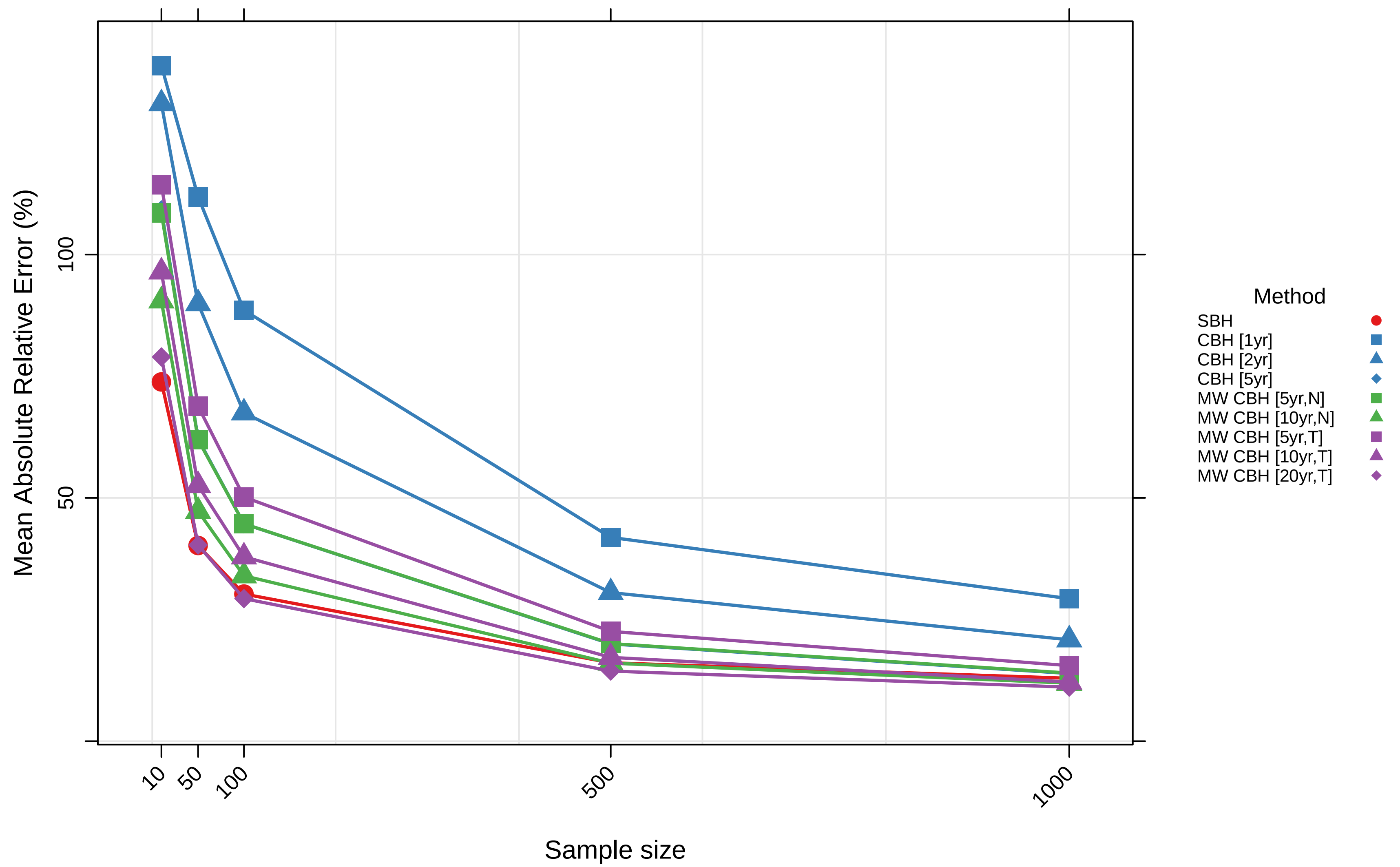

Figure 4: Mean absolute relative error for all methods in the non-stratified analysis. Color of marker indicates the birth history method used, shape distinguishes different variants of each method; SBH=summary birth history; CBH=complete birth history; MW=moving window; N=no weights; T=triangle weights; number in brackets gives period or window length.

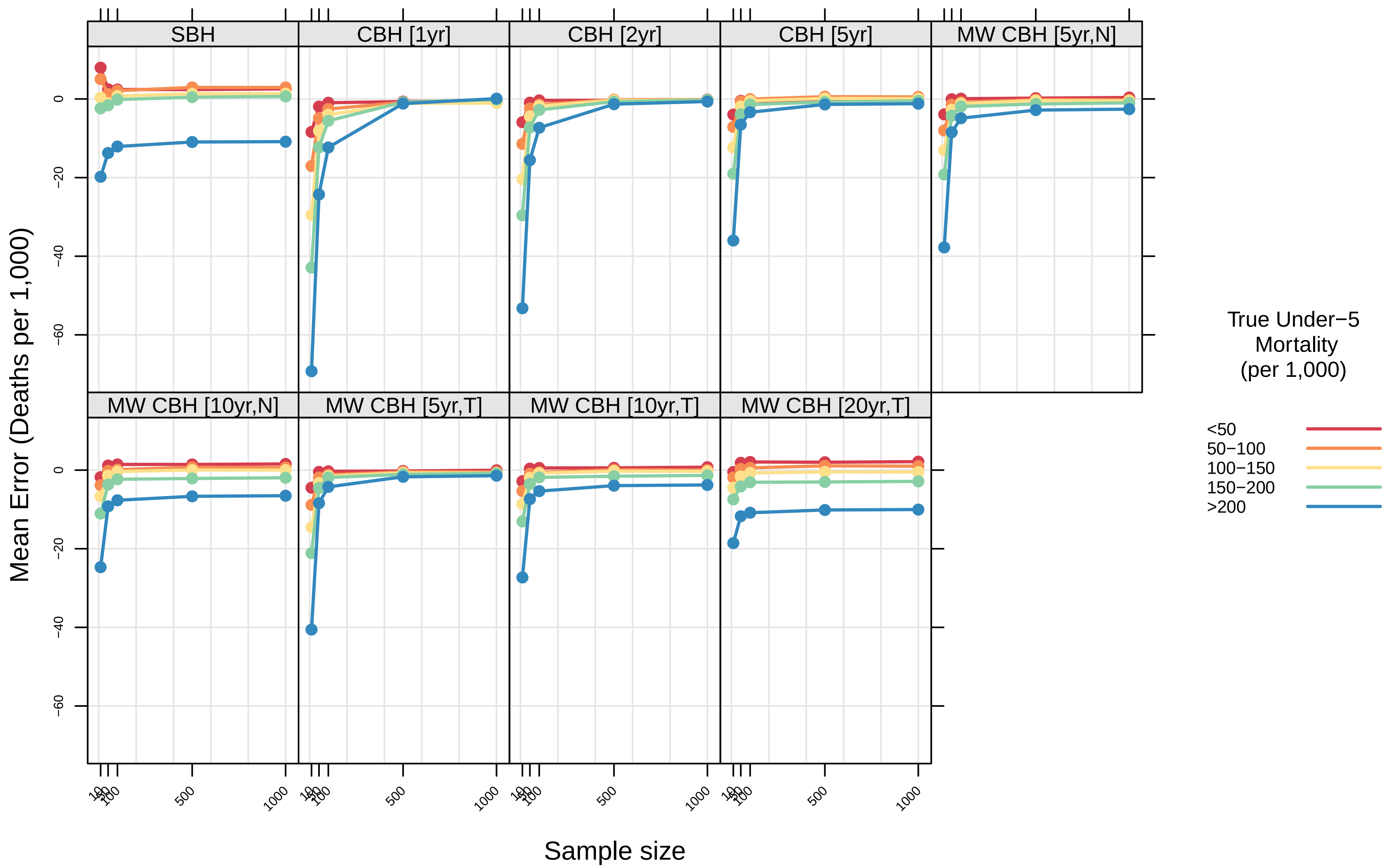

Figure 5: Mean error for all methods in the analysis stratified by true mortality level. Color indicates level of true mortality; each panel shows results for one particular method; SBH=summary birth history; CBH=complete birth history; MW=moving window; N=no weights; T=triangle weights; number in brackets gives period or window length.

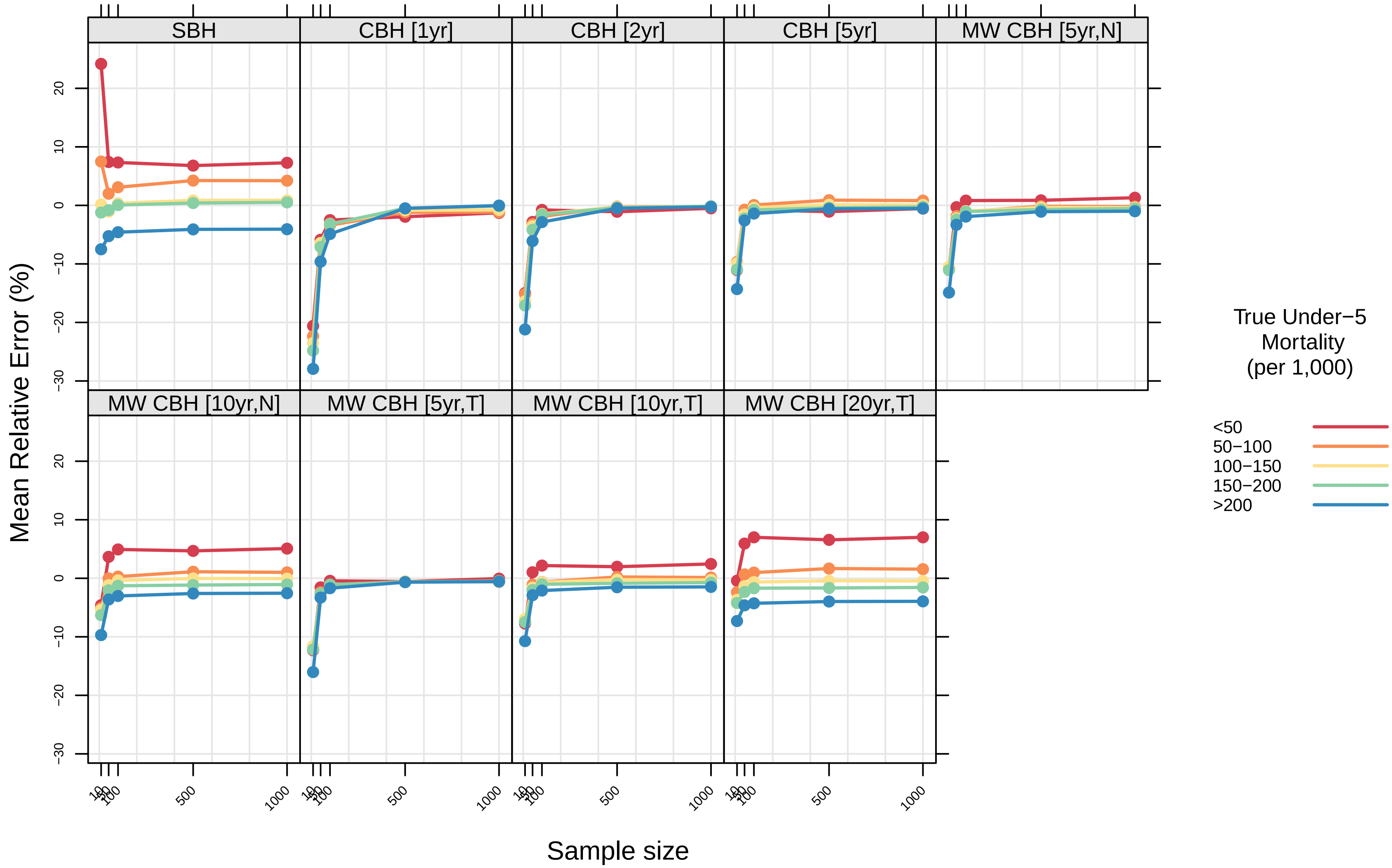

Figure 6: Mean relative error for all methods in the analysis stratified by true mortality level. Color indicates level of true mortality; each panel shows results for one particular method; SBH=summary birth history; CBH=complete birth history; MW=moving window; N=no weights; T=triangle weights; number in brackets gives period or window length.

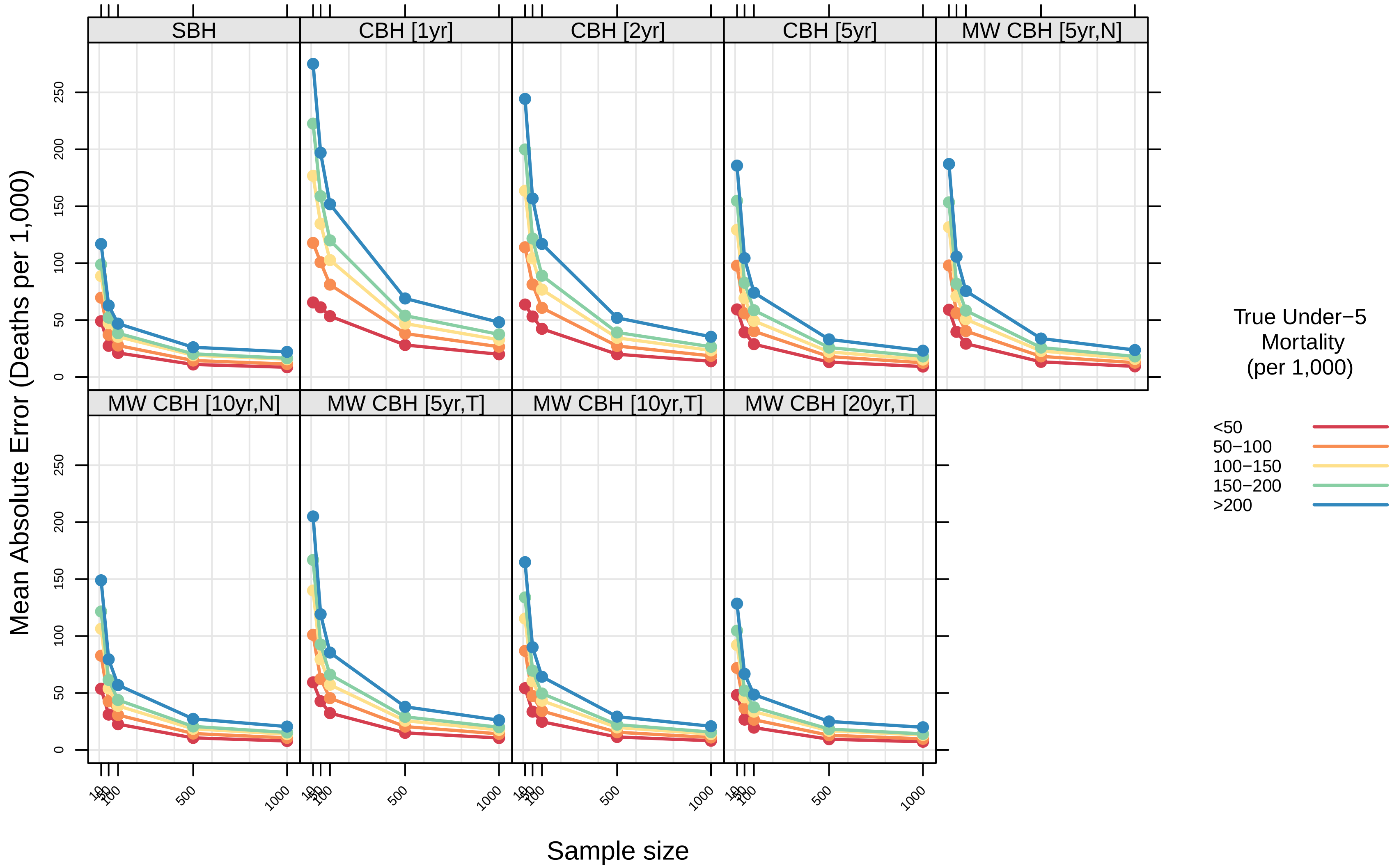

Figure 7: Mean absolute error for all methods in the analysis stratified by true mortality level. Color indicates level of true mortality; each panel shows results for one particular method; SBH=summary birth history; CBH=complete birth history; MW=moving window; N=no weights; T=triangle weights; number in brackets gives period or window length.

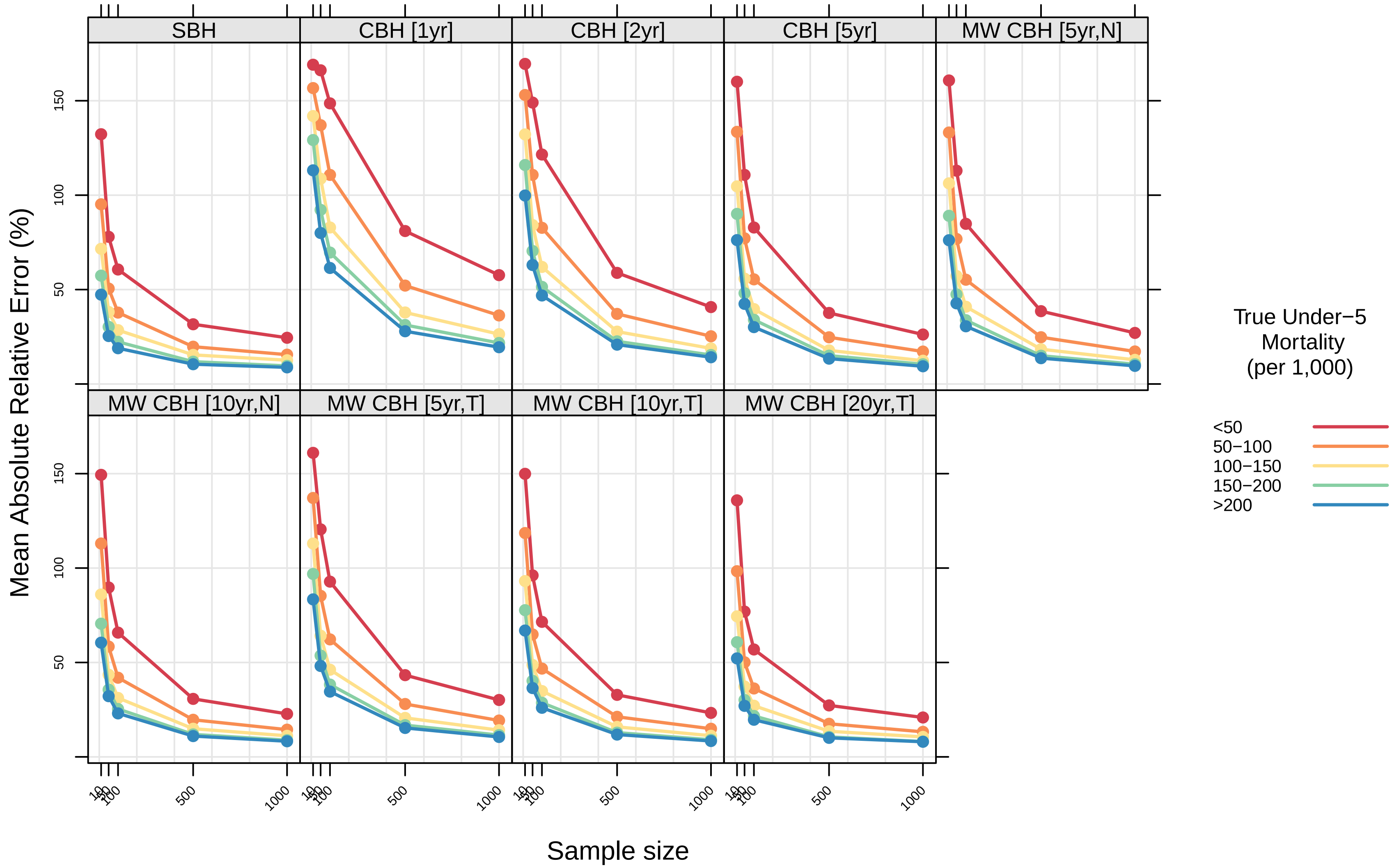

Figure 8: Mean absolute relative error for all methods in the analysis stratified by true mortality level. Color indicates level of true mortality; each panel shows results for one particular method; SBH=summary birth history; CBH=complete birth history; MW=moving window; N=no weights; T=triangle weights; number in brackets gives period or window length.

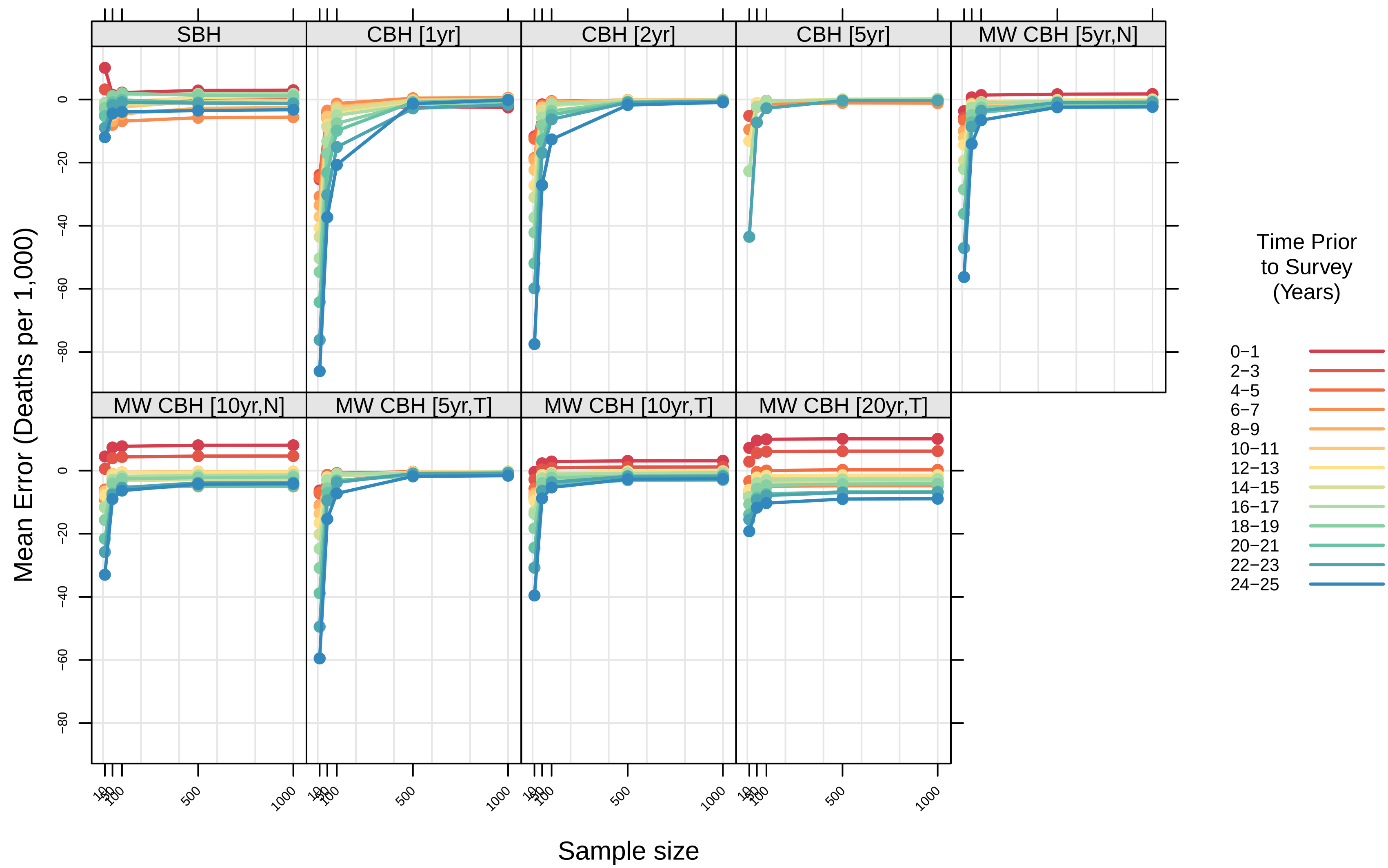

Figure 9: Mean error for all methods in the analysis stratified by time prior to survey. Color indicates time prior to survey; each panel shows results for one particular method; SBH=summary birth history; CBH=complete birth history; MW=moving window; N=no weights; T=triangle weights; number in brackets gives period or window length.

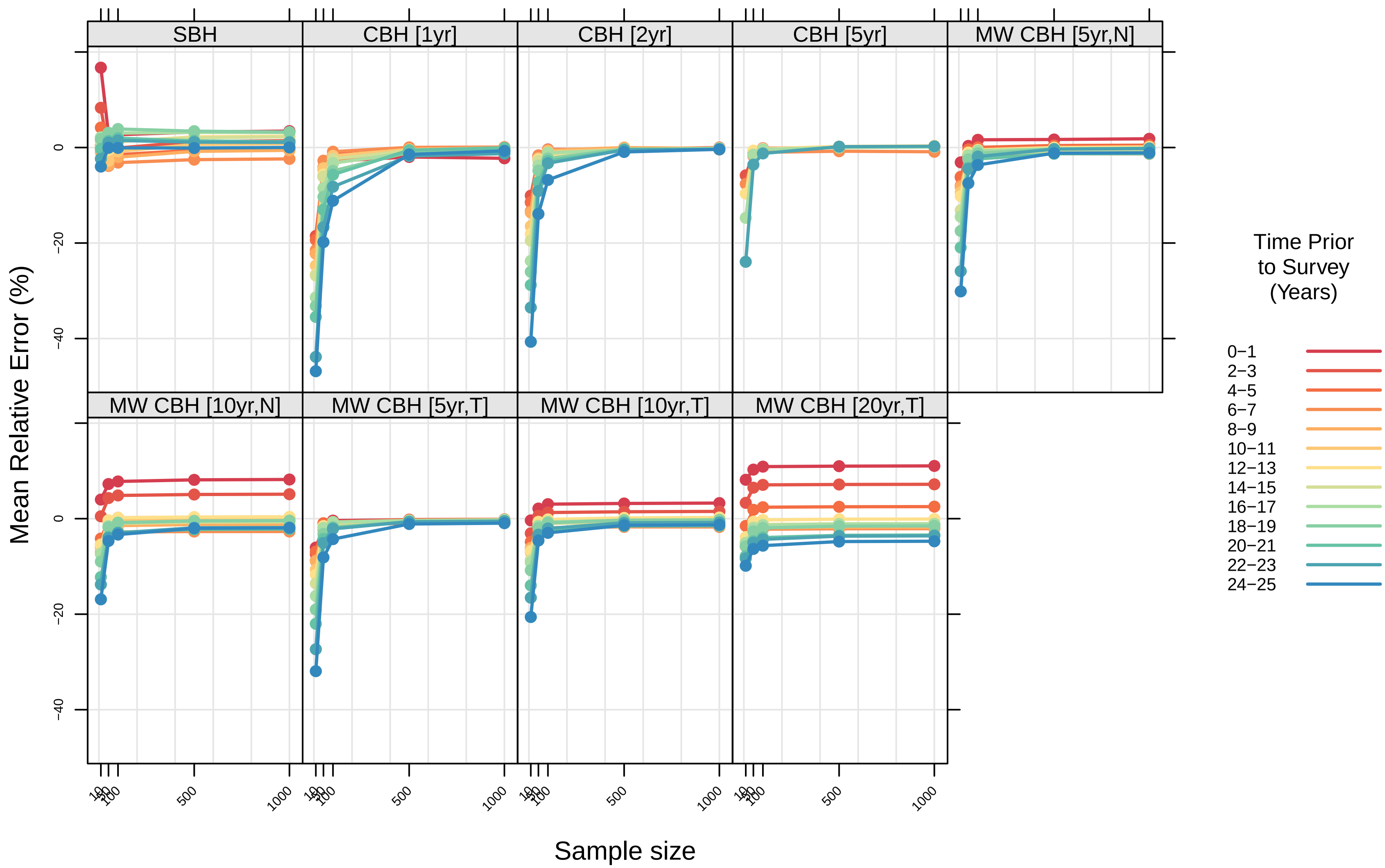

Figure 10: Mean relative error for all methods in the analysis stratified by time prior to survey. Color indicates time prior to survey; each panel shows results for one particular method; SBH=summary birth history; CBH=complete birth history; MW=moving window; N=no weights; T=triangle weights; number in brackets gives period or window length.

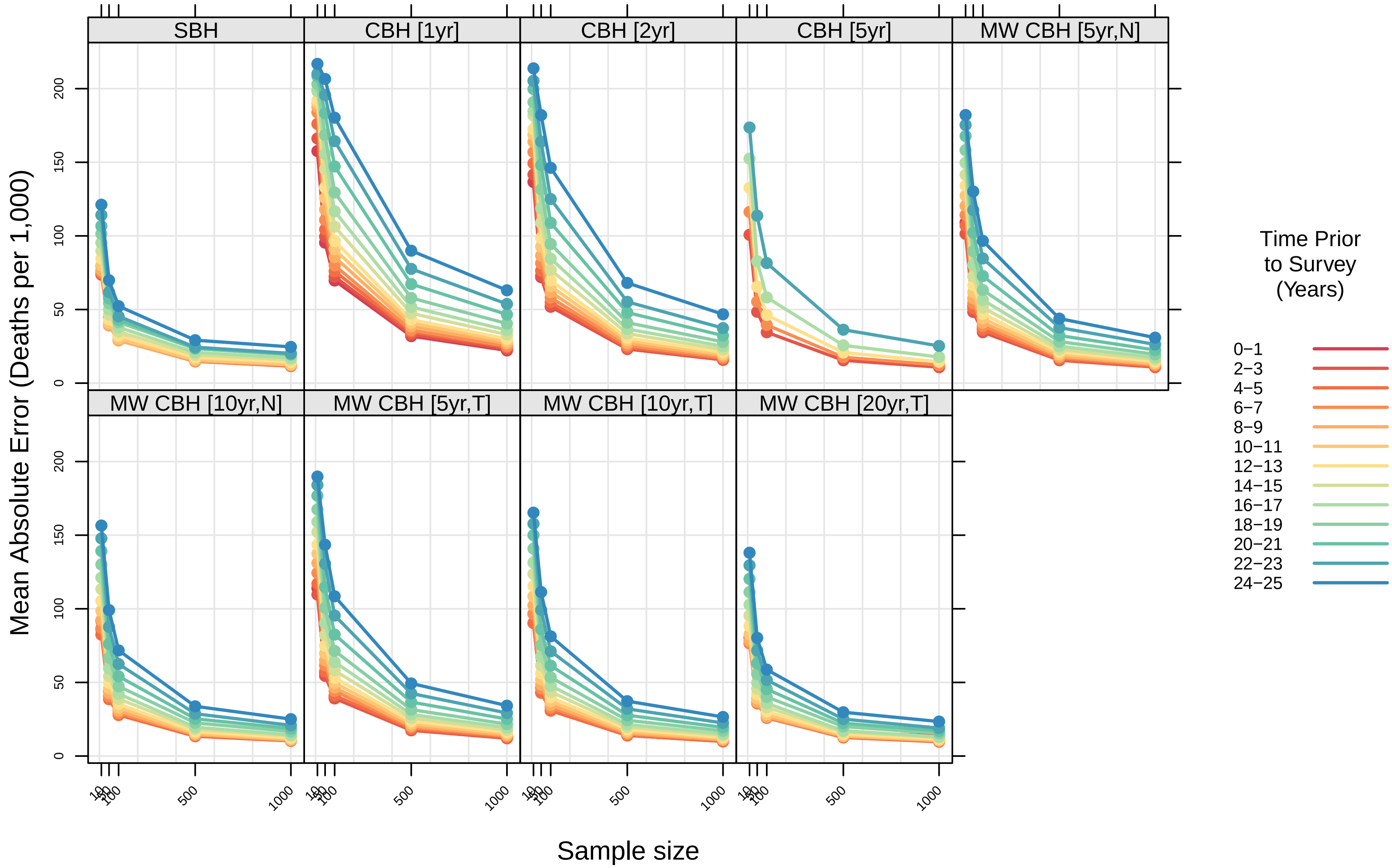

Figure 11: Mean absolute error for all methods in the analysis stratified by time prior to survey. Color indicates time prior to survey; each panel shows results for one particular method; SBH=summary birth history; CBH=complete birth history; MW=moving window; N=no weights; T=triangle weights; number in brackets gives period or window length.

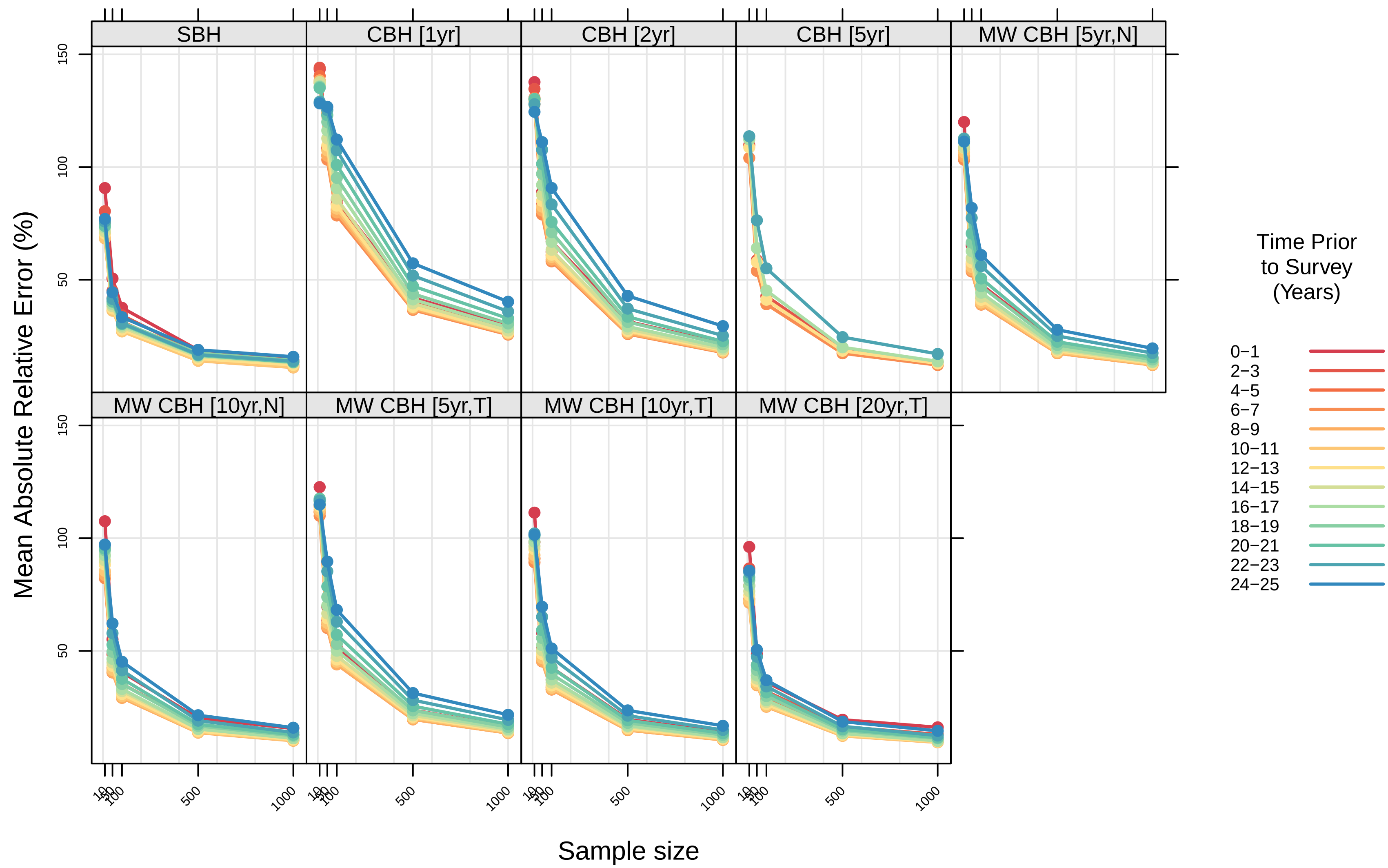

Figure 12: Mean absolute relative error for all methods in the analysis stratified by time prior to survey. Color indicates time prior to survey; each panel shows results for one particular method; SBH=summary birth history; CBH=complete birth history; MW=moving window; N=no weights; T=triangle weights; number in brackets gives period or window length.

References

[1] MEASURE DHS. Demographic and Health Surveys;. Calverton, MD: Macro International, Inc., available: http://www.measuredhs.com.

[2] Rajaratnam J, Tran L, Lopez A, Murray C. Measuring Under-Five Mortality: Validation of New Low-Cost Methods. PLoS Med. 2010;7(4):e1000253.

[3] Lozano R, Wang H, Foreman K, Rajaratnam J, Naghavi M, Marcus J, et al. Progress towards Millennium Development Goals 4 and 5 on maternal and child mortality: an updated systematic analysis. The Lancet. 2011;378(9797):1139–1165.