Title

The boundaries of genocide: Quantifying the uncertainty of the death toll during the Pol Pot regime in Cambodia (1975–79)

Journal
Population Studies, 69(2)

ISSN
0032-4728

Author
Heuveline, Patrick

Publication Date
2015-05-04

DOI
10.1080/00324728.2015.1045546

Peer reviewed

Patrick Heuveline

Department of Sociology

University of California Los Angeles

285 Haines Hall

P.O. Box 951551

Los Angeles, CA 90095-1551

heuveline@soc.ucla.edu

Full Title: **The Boundaries of Genocide: Modeling Credible Intervals for the Death Toll of the Pol Pot Regime (1975-1979)**

Short Title: **Modelling Pol Pot's Death Toll**

Author's name: Patrick Heuveline, PhD

Author's institutional affiliation: California Center for Population Research, University of California Los Angeles

Acknowledgement (Funding)

The author has received no research grant to conduct this research. The research made use of the facilities and resources at the California Center for Population Research, UCLA, which is supported by infrastructure grant R24HD041022 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development.

Abstract

Estimates of excess deaths under Pol Pot's rule of Cambodia (1975-79) range from under one million to over three million. The more plausible among those, methodologically, still vary from one to two million deaths, but this range of independent point estimates has no particular statistical meaning. Stochastically reconstructing population dynamics in Cambodia from extant historical and demographic data yields interpretable distributions of the death toll and other demographic indicators. The resulting 95-percent confidence interval (1.2 to 2.8 million excess deaths) demonstrates substantial uncertainty with regards to the exact scale of mortality, nonetheless excluding nearly half of the extant death-toll estimates. The 1.5 to 2.25 million interval has a greater likelihood (69 percent) of containing the actual number of excess deaths than the wider (one to two million) range of previous plausible estimates. The median value of 1.9 million excess deaths represents 21 percent of the population at risk.

Keywords: Demographic Accounting, Estimation Techniques, Excess Mortality, Violent Deaths, Cambodia, Stochastic, Uncertainty.

Three former high-ranking officials of the Pol Pot regime (PPR thereafter) are currently standing trial in the Extraordinary Chambers in the Courts of Cambodia (ECCC). In accordance with the 1948 United Nations Convention on the Prevention and Punishment of the Crime of Genocide (United Nations 1951), the ECCC is focusing on executions, deaths from exhaustion or starvation related to forced population movements and forced labor, and the treatment of specific ethnic and religious groups. The ECCC is thus reviewing specific actions and policies which took place in the context of a complete overhaul of Cambodia's political, administrative and social structures and the destruction of much of its infrastructure. Curtailing food production and annihilating disease prevention and treatment capacities in particular, the PPR leadership thus contributed to a massive mortality crisis that might have indiscriminately killed more people than any of their more specific, targeted policies.

The total death toll of the short-lived regime (1975–79) has not been definitely established. The authoritative historian of Cambodia, Chandler (1991) writes that the PPR left over a million dead, but extant estimates actually range from 741,000 to 3,315 million. To put those in perspective, the number of global war deaths was recently estimated to average 378,000 annually during the 1985–94 period (Obermeyer et al. 2008), which on the same annualized basis would amount to 1.4 million excess deaths for the entire duration of the PPR. This figure is well within the range of plausible estimates for the PPR death toll alone, while Cambodia's population size then was about .15 percent of the world population size in 1990. The perception that overall mortality increased by such a factor of magnitude partly explains why, though long delayed by domestic and global politics, the ECCC is examining the PPR leadership today, more than three decades after the fall of the regime and more than a decade after the death of Pol Pot himself. If scale considerations are not included in the literal definition of genocide in the 1948

United Nations Convention, they thus remained instrumental in the belated establishment of the ECCC.

In this paper, I seek to assess the credibility of extant estimates of the PPR death toll and to determine how reliably we can estimate the actual level of the mortality under the PPR. In the next section, I begin with a review of extant estimates to try and narrow the field of plausible estimates on methodological and empirical grounds. However, I point out that each of these estimates is but a single "best" point estimate, with an inherent unstated uncertainty that cannot be recovered from published materials. In other words, the plausibility of the range of plausible estimates cannot be readily assessed. To this end, in the rest of the paper I present results of a standard model of population dynamics that incorporates uncertainty in the extant historical and demographic data. In this stochastic framework, the model yields distributions of PPR mortality measures (the death toll, in absolute terms and relative to the population at risk, the number of violent deaths, and life expectancy at birth). As the 1948 United Nations Convention considers among the "acts committed with the intent to destroy, in whole or in part, a national, ethnical, racial or religious group," killing members of a group as well as "imposing measures intended to prevent births," I also estimate the birth deficit under the PPR. These distributions are amenable to a statistical interpretation, and in the final section, I use these distributions to quantify the likelihood of any range of estimate and to compare the plausibility of extant estimates.

Background

Definitions of the death toll

Assessing how many deaths should the PPR be held accountable for involves two distinct types of considerations. The first ones are conceptual: which deaths can be attributed to the PPR? The second ones are methodological: how many of these deaths can be documented? On the conceptual side, the issue concerns possibly non-violent deaths that were nonetheless linked to the PPR policies of population displacement, forced labor, and restricting the availability of food and medicines. If considering only homicides perpetrated by PPR cadres or even other violent deaths (e.g., fatalities or mine-related accidents) seems insufficient, using the total number of deaths is equally unsatisfactory. Few opinions to the contrary (Seltzer 2008), the general agreement is that no mortality is not a tenable benchmark for assessing a government's responsibility, since deaths remain unavoidable under the best government and circumstances.

The common approach in causal analysis involves comparing the prevailing situation to a counterfactual one; in this case the actual number of deaths to a counterfactual number of deaths had the PPR never taken over. Less straightforward is the definition of the reference period for comparing the actual and the counterfactual numbers of deaths. Should a regime taking over after years of civil war or a government leaving power in the midst of a famine, to which its disastrous resource management contributed, be accountable for some of these war-related or famine deaths? Even though these deaths might occur while the regime is not yet or no longer in power, the answer is probably yes, or at least in part. In practice however, a consensus would seem beyond reach as far as delineating the beginning and end point for this period of accountability and what exact share of these deaths should be attributed to one regime rather than the ones that preceded or followed. In spite of the potential limitations, we follow most previous analysts and focus on the PPR death toll, defined as the difference in the estimated number of deaths between prevailing and counterfactual mortality—that is, "excess" mortality—strictly during the period the PPR was in power. We also discuss the estimation of the subset of these excess deaths that can be identified as violent deaths.

A typology of estimation approaches

The different approaches developed to estimate the death toll have been reviewed elsewhere (Heuveline 2001) and are only briefly described here, with further details provided in Appendix I. For expositional clarity, these approaches can be classified based on the type of evidence they most extensively rely on, even though most combine evidence from more than a single type: verbal (testimonies on the death/survival of relatives), demographic (censuses and surveys), or physical (bone remains). First available chronologically, verbal evidence was collected from refugees who had fled Cambodia toward the end of the PPR (Central Intelligence Agency 1980; Ea 1981) before a few independent researchers were able to collect testimonies in Cambodia as well (Kiernan 1996; Heder 1997). Similar testimonies continued to be collected into the early 1990s, taking advantage of the gradual accessibility of more remote Cambodian provinces (Sliwinski 1995). In the meantime, the post-PPR government had produced its own estimate by attempting to gather an exhaustive testimonial account throughout the country (Gordon 2007).

Even though this first approach may require some demographic data to scale up survival probabilities from the sample to the whole population, the second approach can be characterized by its near-exclusive use of demographic evidence. This approach yields an estimate of excess deaths as a residual (unaccounted for) term in a demographic reconstruction with actual population change, actual fertility and migration, and counterfactual mortality. The residual estimate can be obtained from a simple accounting equation with aggregate numbers of births, counterfactual deaths and migrants, or preferably by performing this accounting for each gender and age groups separately (Preston et al. 2001). Whereas samples of testimonies can be collected relatively quickly, however, a scientific population census takes years to organize, even in peace time. The first estimates of excess deaths both by aggregate accounting (Vickery 1984) and disaggregated by gender and age groups (Banister and Johnson 1993) were based instead on official figures. These were obtained from a simple tally of head counts collected by the local administration in 1980 and updated annually thereafter. Disaggregated accounting was further refined as new demographic data became available: the 1992 electoral rosters assembled by the United-Nations (Heuveline 1998), the first post-PPR General Population Census in 1998 (Neupert and Prum 2005), and the most recent one to date in 2008 (Lognard 2012).

The last type of estimation, based on physical evidence (e.g., bone remains), only concerns violent deaths. Beginning with a full mapping of mass grave pits conducted from satellite imagery, remote sensing, and ground penetrating radars, this approach continues with interviews, and in some cases, forensics to determine the time and cause of each death (Etcheson 2005). Violent deaths had previously been estimated from testimonies using respondents' reported cause of death, and apportioning the number of excess deaths between violent and non-violent (Ea 1981; Sliwinski 1995; Heder 1997). Disaggregated accounting by gender and age group has also been used to estimate the number of violent deaths (Heuveline 1998). This estimation requires a comparison with the counterfactual increase in non-violent mortality only which can be derived from life tables for high-mortality populations (Preston et al. 1993).

An assessment of extant estimates

As summarized in Table 1, the seven extant estimates of violent deaths are widely scattered. The lowest three are in the order of 100,000 to 300,000 and the highest two almost identical at 1.1 million. The nine extant estimates of excess deaths range from 741,000 to 3.315 million. The highest figure is the governmental estimate and also the only one based on a direct tally of excess-death reports. The estimate is typically dismissed as propaganda on the ground that no attempt was made to identify multiple counts or to adjust for the differential probabilities of being reported. Likewise, the lowest figure originates in the only total accounting not disaggregated by gender and age groups. These two idiosyncratic estimations aside, the remaining estimates of the PPR death toll range from 1.0 million to 2.1 million.

Coincidently or not, the one-to-two-million interval appears to be the most often reported range of PPR death toll estimates. The ECCC, historians, and the general public would likely prefer a single, consensual estimate, but one could hardly adjudicate between these figures derived with such different methodologies and evidence bases. For each type of evidence, a trend toward higher estimates over time can be detected in Table 1, but the later estimates are not necessarily superior to the earlier ones. Collected later, testimonies can originate from a more diverse set of respondents, but require them to remember more distant events. This longer recall period may reduce data quality. For demographic data, the 1998 census is clearly a more reliable source than previous surveys or administrative counts. To estimate the desired population size at the end of the PPR, however, the reconstruction then requires assumptions about fertility, mortality and migration for nearly two decades after the fall of the PPR (1979–98).

Rather than statistically interpretable, the range of extant estimates is merely a collection of point estimates, each one with its own uncertainty, unstated and unrecoverable ex-post from the published documentation. One particular sample-based estimation yielded a 1.843-to-1.869-million interval (Sliwinski 1995) that merely corresponds to two alternate assumptions, since a full assessment of uncertainty is impossible. First, the classic sampling designs with readily calculated standard errors cannot be implemented in such crisis settings (Hagan and Palloni 2005). Second, the conditions under which direct mortality estimates from sibling data would be unbiased (Gakidou and King 2003) were not met either. As expected, in particular from the PPR policies of targeting potential enemies along kinship lines (Kiernan 1996; Weitz 2003), mortality clustering within families has been identified (Sliwinski 1995). However, available methods to adjust for the differential probabilities of being reported (Trussel and Rodriguez 1990) were not used. Geographical clustering, uncertainty about the total population size to which to prorate estimated survival/death ratios, and uncertainty about baseline mortality add potential biases and uncertainty to the sample-based estimates.

One accounting exercise provides a relatively wide range rather than a single number (Heuveline 1998), but the high and low end of the range still just correspond to alternate assumptions about fixed parameters. Lee (1985, p.234) has long pointed out that any single reconstruction represents only one plausible (e.g., internally consistent) demographic trajectory "among an infinity of different but equally plausible ones." The quantities that must exactly satisfy the demographic accounting relationships cannot be measured without errors (Wheldon et al. 2013), however, or more importantly here, known without much uncertainty. Each estimate of excess deaths—a residual term in the balancing equation of population change—is then just one of the estimates corresponding to one set of plausible estimates of other parameters in the demographic reconstruction. However, one of the advantages of the demographic reconstruction approach is precisely that it can be used to formally investigate this uncertainty.

Methods

Demographic reconstruction and models of population change

The logic of demographic reconstruction rests on the balancing equation of population change

(**1**): $N(t) = N(t\text{-}x) + B(t\text{-}x,t) – D(t\text{-}x,t) + I(t\text{-}x,t) – O(t\text{-}x,t)$

where $N(t)$ is the number of persons alive in the population at time t, and $B(t\text{-}x,t)$, $D(t\text{-}x,t)$, $I(t\text{-}x,t)$, and $O(t\text{-}x,t)$ are respectively the number of births, deaths, in-migrants, and out-migrants between time $t\text{-}x$ and time t (Preston et al. 2001). When only some of the components in Equation 1 are known, the equation can be used to derive the sum of the remaining components. This is typically used to the estimate the net flow of in and out migration, which is often poorly recorded, from the number of births and deaths, which are recorded in vital statistics, and population sizes from two population censuses (Shryock and Siegel 1973). In a similar vein, the approach could be used to estimate instead the number of deaths in the interval between two population counts from estimates of the number of births, in-migrants and out-migrants in the same interval. The estimation of the death toll further distinguishes excess mortality from (counterfactual) baseline mortality. Assuming that we can estimate the counterfactual number of deaths corresponding to baseline mortality between time $t\text{-}x$ and time t, $D^{C*}(t\text{-}x,t)$, the number of excess deaths between time $t\text{-}x$ and time t, $D^{E*}(t\text{-}x,t)$, can then be derived as

(**2**): $$D^{E*}(t\text{-}x,t) = N^*(t\text{-}x) - N^*(t) + B^*(t\text{-}x,t) - D^{C*}(t\text{-}x,t) + I^*(t\text{-}x,t) - O^*(t\text{-}x,t)$$

To focus on violent mortality only, we need to estimate instead the counterfactual number of non-violent deaths corresponding to the actual increase in non-violent mortality only between time $t\text{-}x$ and time t, $D^{N*}(t\text{-}x,t)$, and derive the number of violent deaths between time $t\text{-}x$ and time t, $D^{V*}(t\text{-}x,t)$, as

(**3**): $$D^{V*}(t\text{-}x,t) = N^*(t\text{-}x) - N^*(t) + B^*(t\text{-}x,t) - D^{N*}(t\text{-}x,t) + I^*(t\text{-}x,t) - O^*(t\text{-}x,t)$$

Likewise, this approach can be used to estimate the birth deficit, that is, the difference between the counterfactual, expected number of births expected had fertility remained at its baseline level and the actual number of births. From an estimate of the counterfactual number of births corresponding to baseline fertility between time t-x and time t, $B^{C*}(t\text{-}x,t)$, the number of missing births between time t-x and time t, $B^{M*}(t\text{-}x,t)$, can then be derived as

(**4**): $$B^{M*}(t\text{-}x,t) = N^*(t\text{-}x) - N^*(t) + B^C{}^*(t\text{-}x,t) - D^*(t\text{-}x,t) + I^*(t\text{-}x,t) - O^*(t\text{-}x,t)$$

Like demographic projections, however, demographic reconstruction is preferably carried on with a model of population dynamics disaggregated by gender and age-group rather than from population-level counts. Disaggregation presents at least two advantages. First, this approach uses information on gender and age distribution from census data and takes advantage of the extensive variation in survival over the life course which is embodied in age patterns of mortality models (Heuveline and Clark 2011). Better grounded empirically, the estimation of the counterfactual number of deaths (or births) is improved, and in turn, so is the estimation of excess deaths (or the birth deficit). Second, the breakdown of excess mortality by gender and age group itself provides useful information. Comparison of the age distribution of excess mortality and of the age distribution that could be obtained by modelling an increase in mortality from natural causes with high-mortality life tables (Preston et al. 1993) allows for a decomposition of excess mortality between violent and non-violent mortality (Heuveline 1998).

The cohort-component model of population projection (CCMPC thereafter) is by far the most commonly used model of population dynamics (Preston et al. 2001). In this model, the number of persons alive in the population at time t, $N(t)$, is broken down by gender and age groups of the same length x, except for the oldest persons of each gender, who are grouped in open-ended age intervals. The CCMPC can more compactly be written in matrix notation as

(**5**): $$\boldsymbol{N}(t) = [\boldsymbol{L}(\boldsymbol{F}(t\text{-}x,t),\beta(t\text{-}x,t),\boldsymbol{S}(t\text{-}x,t))^*\{\boldsymbol{N}(t\text{-}x)+(\alpha(t\text{-}x,t).\boldsymbol{\Delta}(t\text{-}x,t))\}]+[(1\text{-}\alpha(t\text{-}x,t)).\ \boldsymbol{\Delta}(t\text{-}x,t)]$$

where $\boldsymbol{N}(t)$ is the vector of gender and age-groups at time t, $\boldsymbol{\Delta}(t\text{-}x,t)$ is the vector of differences between the number of in- and out-migrants between times $t\text{-}x$ and t by gender and age-group, and $\boldsymbol{L}(F(t\text{-}x,t),\beta(t\text{-}x,t),S(t\text{-}x,t))$ is a matrix function of the vector of female age-specific fertility, $\boldsymbol{F}(t\text{-}x,t)$, the sex-ratio at birth scalar, $\beta(t\text{-}x,t)$, and the vector of age-specific survival ratios, $\boldsymbol{S}(t\text{-}x,t)$, all of them between times $t\text{-}x$ and t as well. The value of parameter $\alpha(t\text{-}x,t)$ can be chosen between 0 and 1 depending on the flow of migration between times $t\text{-}x$ and t, with .5 representing an even flow (constant number of migrants per unit of time) and 1 representing an "instantaneous" migration concentrated at time $t\text{-}x$. Implicit in Equation **5** is the common assumption that once they enter the population, immigrants survive to time t in the same proportion as people already in the population at time $t\text{-}x$.

The logic used to derive Equations **2** to **4** from Equation **1** is readily repeated to derive excess mortality, violent mortality, or missing births from Equation **5**. In the case of excess deaths, they can be estimated from

$$(\mathbf{6}):\quad \boldsymbol{N}(t) = [\boldsymbol{L}(\boldsymbol{F}(t\text{-}x,t),\beta(t\text{-}x,t),\boldsymbol{S}^{C}(t\text{-}x,t))*\{\boldsymbol{N}(t\text{-}x)+(\alpha(t\text{-}x,t).\boldsymbol{\Delta}(t\text{-}x,t))+(\varepsilon(t\text{-}x,t).\boldsymbol{D}^{E}(t\text{-}x,t))\}]$$
$$+ [\{(1\text{-}\alpha(t\text{-}x,t)).\ \boldsymbol{\Delta}(t\text{-}x,t)\}+\{(1\text{-}\varepsilon(t\text{-}x,t)).\boldsymbol{D}^{E}(t\text{-}x,t)\}]$$

where $\boldsymbol{S}^{C}(t\text{-}x,t)$ is now the counterfactual survival ratio corresponding to baseline mortality between times $t\text{-}x$ and t, $\boldsymbol{D}^{E}(t\text{-}x,t)$ is the vector of excess deaths times $t\text{-}x$ and t by gender and age groups, and $\varepsilon(t\text{-}x,t)$ is a scalar whose value between 0 and 1 accounts for the timing of excess mortality between times $t\text{-}x$ and t. The solution now takes the form

$$(\mathbf{7}):\quad \boldsymbol{D}^{E}\boldsymbol{*}(t\text{-}x,t) = \varepsilon^*(t\text{-}x,t).[\{\boldsymbol{L}'(\boldsymbol{S}^{C}\boldsymbol{*}(t\text{-}x,t))\boldsymbol{*}(\boldsymbol{N}\boldsymbol{*}(t)\text{-}((1\text{-}\alpha^*(t\text{-}x,t)).\boldsymbol{\Delta}\boldsymbol{*}(t\text{-}x,t)))\}$$
$$\text{-}\{\boldsymbol{N}\boldsymbol{*}(t\text{-}x)+(\alpha^*(t\text{-}x,t).\boldsymbol{\Delta}\boldsymbol{*}(t\text{-}x,t))\}]$$
$$+\ (1\text{-}\ \varepsilon^*(t\text{-}x,t)).[\{\boldsymbol{N}\boldsymbol{*}(t)\text{-}((1\text{-}\alpha^*(t\text{-}x,t)).\boldsymbol{\Delta}\boldsymbol{*}(t\text{-}x,t))\}$$

$$-\{\boldsymbol{L}(\boldsymbol{F}^*(t\text{-}x,t),\beta^*(t\text{-}x,t),\boldsymbol{S^C}^*(t\text{-}x,t))^*(\boldsymbol{N}^*(t\text{-}x)+(\alpha^*(t\text{-}x,t).\boldsymbol{\Delta}^*(t\text{-}x,t)))\}]$$

where $\alpha^*(t\text{-}x,t)$ and $\varepsilon^*(t\text{-}x,t)$ are scalars that relate to the scalars $\alpha(t\text{-}x,t)$ and $\varepsilon(t\text{-}x,t)$ respectively, and $\boldsymbol{L'}(\boldsymbol{S^C}^*(t\text{-}x,t))$ is a matrix such that $\boldsymbol{L}(\boldsymbol{F}^*(t\text{-}x,t),\beta^*(t\text{-}x,t), \boldsymbol{S^C}^*(t\text{-}x,t))^* \boldsymbol{L'}(\boldsymbol{S^C}^*(t\text{-}x,t))^* \boldsymbol{N}^*(t)$ equals $\boldsymbol{N}^*(t)$. $\boldsymbol{L'}(\boldsymbol{S^C}^*(t\text{-}x,t))$ is not entirely determined because the oldest closed age interval and the open age interval at time $t\text{-}x$ both contribute to the closed age interval at time t, and multiple combinations of the survival ratios for these two age groups are possible. An additional constraint needs to be applied to select a particular $\boldsymbol{L'}(\boldsymbol{S^C}^*(t\text{-}x,t))$ (see Lee (1985) for a fuller discussion), but for relatively short reconstruction periods, the results are not too sensitive to this particular step. More importantly, Equation **7** emphasizes the potential impact of the timing of excess mortality as it relates to the corresponding value of $\varepsilon^*(t\text{-}x,t)$, which act as a weighting function between two terms. These two terms represent the difference between the counterfactual and the actual population size at time $t\text{-}x$ and time t respectively, which can be very different when population grows or declines substantially between time $t\text{-}x$ and time t.

Since the death toll is here defined as the number of excess deaths during the period the country was ruled by the PPR, we can use Equation **7** to deterministically estimate the death toll by setting $t\text{-}x$ as the onset of the regime, April 17th, 1975, and t as the day it fell, January 7th, 1979. Population counts are not available for these dates, so they must also be derived from population size estimates at the closest available date and estimates of fertility, survival and migration in-between. Formally, Equation **7** must be completed with

(**8**): $$\boldsymbol{N}^*(t\text{-}x) = [\boldsymbol{L}(\boldsymbol{F}^*(t_1,t\text{-}x),\beta^*(t_1,t\text{-}x),\boldsymbol{S}^*(t_1,t\text{-}x))^*\{\boldsymbol{N}^*(t_1)+(\alpha^*(t_1,t\text{-}x).\boldsymbol{\Delta}^*(t_1,t\text{-}x))\}] + [(1\text{-}\alpha^*(t_1,t\text{-}x)).\ \boldsymbol{\Delta}^*(t_1,t\text{-}x)]$$

and (**9**): $$\boldsymbol{N}^*(t) = [\boldsymbol{L}'(\boldsymbol{S}^*(t,t_2))^*\{\boldsymbol{N}^*(t_2)\text{-}((1\text{-}\alpha^*(t,t_2)).\boldsymbol{\Delta}^*(t,t_2))\}] - \{\alpha^*(t,t_2).\boldsymbol{\Delta}^*(t,t_2)\}]$$

In this case, we can set t_1 as April 18th, 1962, date of the last pre-PPR census, and t_2 as March 2nd, 1998, date of the first post-PPR census.

Instead of an estimate of the total death toll, an estimate of the number of violent deaths is derived by substituting $\boldsymbol{S}^{N}*(t\text{-}x,t)$, instead of $\boldsymbol{S}^{C}*(t\text{-}x,t)$ in Equation **7**, where $\boldsymbol{S}^{N}*(t\text{-}x,t)$ now represents the survival ratios estimated to correspond to a counterfactual increase in only non-violent mortality. Finally, a birth deficit can also be estimated by comparing the actual number of births to a counterfactual number. The actual number, $B(t\text{-}x,t)$, is a function of the vector of female age-specific fertility rates, $\boldsymbol{F}(t\text{-}x,t)$, and the vector of person-years of exposure in the different age-groups of the female population between times $t\text{-}x$ and t, $\boldsymbol{P}(t\text{-}x,t)$. With the conventional approximation of person-years (Preston et al. 2001), the number of missing births is simply estimated as

(**10**): $\boldsymbol{B}^{M}*(t\text{-}x,t) = (x/2) * \{\boldsymbol{F}^{C}*(t\text{-}x,t) - \boldsymbol{F}*(t\text{-}x,t)\} * \{\boldsymbol{N}*(t\text{-}x) + \boldsymbol{N}*(t)\}$

where $\boldsymbol{F}^{C}*(t\text{-}x,t)$ is the vector of counterfactual fertility rates.

We could estimate the counterfactual number of births had both fertility and mortality remained at their baseline level. As shown in Equation **10**, however, we rather estimate the counterfactual number of births had fertility only remained at its baseline level (i.e., using prevailing mortality). While this is a more restrictive definition of the birth deficit, the approach has the advantage of being additive in the sense that the death toll and the birth deficit can be added to estimate how much larger would the population size be with counterfactual fertility and mortality both at baseline.

We could similarly define the impact on population size of the massive emigration of Cambodian refugees induced by the PPR. However, it is of little interest under our operative definitions, which limit the estimation of the PPR impact strictly to the PPR era. Leaving Cambodia during the PPR was extremely difficult and the bulk of the Cambodian refugees left either just before the PPR or, in much higher numbers, in the years after the PPR.

A formal sensitivity analysis using the CCMPC

Equations **7** to **10** provide estimates of the number of excess deaths and the number of missing births, from single values (point estimates) of $\boldsymbol{N}(t_1)$, $\boldsymbol{S}(t_1,t\text{-}x)$, $\boldsymbol{F}(t_1,t\text{-}x)$, $\beta(t_1,t\text{-}x)$, $\alpha(t_1,t\text{-}x)$, $\boldsymbol{\Delta}(t_1,t\text{-}x)$, $\boldsymbol{S}^C(t\text{-}x,t)$, $\boldsymbol{F}^C(t\text{-}x,t)$, $\boldsymbol{F}(t\text{-}x,t)$, $\beta(t\text{-}x,t)$, $\alpha(t\text{-}x,t)$, $\boldsymbol{\Delta}(t\text{-}x,t)$, $\varepsilon(t\text{-}x,t)$, $\boldsymbol{S}(t,t_2)$, $\alpha(t,t_2)$, and $\boldsymbol{\Delta}(t,t_2)$. Uncertainty in any of these point estimates should translate into uncertainty in the estimates of the death toll and the birth deficit. We can implement the model so to provide estimates of this uncertainty as well, by treating the parameters of the CCMPC reconstruction as random within their respective distributions rather than fixed.

Since the logic of the estimation is identical, we can return to the balancing equation of population change (Equation **1**) to discuss this approach. If we have independent estimates of $N^*(t\text{-}x)$, $N^*(t)$, $B^*(t\text{-}x,t)$, $D^*(t\text{-}x,t)$, $I^*(t\text{-}x,t)$, and $O^*(t\text{-}x,t)$, these estimates might not satisfy Equation **1**, a discrepancy known as the “error of closure”(Preston et al. 2001). The discrepancy arises from measurement errors that affect the estimation of the different components of Equation **1**, and the error of closure must somehow be “apportioned” to the measurement errors affecting each of the different components. The apportioning process becomes relatively complex in a full CCPMC reconstruction covering a long period and data are disaggregated by gender and age groups. Wheldon et al. (2013) have developed an elegant solution to the estimation of uncertainty in the demographic reconstruction when the main source of uncertainty is measurement error. The solution depends on the variance structures of the various parameters.

While their simplifying assumption of constant variance across age and time for each demographic parameter is plausible for most settings, it clearly cannot be applied here. The assumption could be modified, but an alternative is not easily formulated, especially for the PPR period estimates.

A different approach is to use computer simulations to formally investigate the uncertainty in the reconstruction through a systematic assessment of the sensitivity of its outputs to measurement error in its input parameters (Iman and Helton 1988). This requires that extant information be used first to derive not just point estimates of the scalars and vectors in Equations **7** to **10**, but rather to estimate their full distributions. Specifically, we identify 47 critical parameters in the following implementation of the CCMPC reconstruction. First, we structure the vectors in Equations **7** to **10** in 5-year age groups. This is a convenient value because mortality models provide survival ratios for 5-year age groups, and the use of these models greatly reduces the dimension parameter space. This is also acceptable because mortality varies relatively regularly within 5-year age groups, making the age distribution within 5-year age groups relatively unimportant. This is only acceptable, however, as long as time trends are also relatively continuous over 5-year time intervals. In this case, discontinuities in demographic trends were associated with the following political events: the March-18th-1970 coup which ushered in a civil war, the PPR takeover on April 17th, 1975, and the PPR fall on January 7th, 1979. The forward estimation of the April 17th, 1975 population characteristics from the 1962 census data thus involves three distinct reconstruction intervals (1962–67, 1967–70, and 1970–75). Similarly, the backward estimation of the January 7th, 1979 population characteristics from the 1998 census data involves four distinct reconstruction intervals (1979–83, 1983–88, 1988–93, and 1993–98).

We then specify a separate probability distribution for each of 47 critical parameters of the reconstruction, as shown in Appendix Table A1. Of these 47 parameters, 18 are used from the forward reconstruction from the time of the 1962 census to the time of the 1975 PPR take-over, 24 are used for the backward reconstruction from the time of the 1998 census to the time of the 1979 PPR fall, and 5 are used to reconstruct the PPR period (by forward and backward iterations). Each distribution is set based on a review of data sources that have been available to and used in some or all of the extant estimations, except for only recently available data on fertility during the PPR (Heuveline and Poch 2007). Extant estimations have relied on choosing one source over the others, however, whereas here we can consider all these sometimes discordant sources jointly in defining our parametric distributions. (Supplementary Materials on the empirical justification of the distributions are available from the author.) We are careful to choose reconstruction parameters that do not clearly violate the assumption that these parameters are independent from one another. For instance, to set mortality levels in two consecutive periods, which are likely to be correlated, we chose as our parameters the mortality level in the first period and the mortality gains from one period to the next.

To represent the effect of the joint variation of the input parameters, each within its specified probability distribution, we then use the Monte Carlo simulation technique. Namely, we pick one input parameter value at a time, by taking the value corresponding to a randomly selected percentile in distribution between .5 and 99.5 (with increments of one unit). We then combine the input parameter values in the CCMPC reconstruction to obtain one set of output values. Each of 100 forward simulations of the 1975 population by age and sex is paired with each of 100 backward simulations of the 1979 population by age and sex, and for each of the 10,000 such pairs, we randomly draw a distinct set of five 1975-79 parameters. From these multiple draws, we record the values of the demographic indicators from each reconstruction to obtain frequency distributions, with the usual summary statistics such as the mean, median, standard deviation (SD) and confidence intervals (CI). Because making additional draws and simulations is relatively straightforward, we use simple random sampling and a large number of draws rather than alternate sampling designs that might be more efficient but more difficult to implement , e.g., Latin Hypercube sampling based on random sampling without replacement (Handcock 1989).

Results

Figure 1 shows our probabilistic estimates of the total population size between the time of the first census ever conducted in Cambodia and the time of the first post-PPR census (1962–98). Uncertainty about the total population size increases from 1962 to 1975, as it compounds uncertainty from the various demographic parameters needed to project the population forward from the time of the 1962 census. This is most noticeable after 1970, as parameter uncertainty is even greater during the civil war leading to the PPR takeover (1970–75) than in earlier periods. Uncertainty with respect to the total population size peaks at the onset of the PPR, and the range of our 95% CI is over one million, from 7.62 to 8.75 million (median 8.11 million.) Uncertainty also increases as we proceed backward from the 1998 census toward the end of the PPR.

Table 2 focuses on the demographic estimates of the PPR period. Our median estimate of the number of violent deaths is 1.09 million, over 80% of these deaths being male (884,000). With a SD of 186,000, the 95% CI spreads from 728,000 to 1.455 million, however. As regards the total PPR death-toll, our median estimate is 1.90 million excess deaths, two thirds of these excess deaths being male (1.27 million). The corresponding SD is 374,000, and the 95% CI spreads from 1.20 to 2.76 million. Both median values are thus within the range of plausible extant estimates reviewed in Table 1.

The full distribution of death toll estimates is displayed in Figure 2. The "standard" one-to-two-million interval only appears to have a 60.19% likelihood of containing the actual value, with the following breakdown: 12.83% (1.0-1.5 million), 20.43% (1.5-1.75 million), and 26.93% (1.75-2.0 million). As the likelihood is actually higher between 2.0 and 2.25 million (21.57%) than between 1.0 and 1.5 million, the 1.5-2.25 million range actually has a higher likelihood of including the actual value (68.93%) than the wider 1.0-2.0 million range.

Estimates of the PRR death toll are often expressed in relative terms, as a ratio to the population size at the outset of the PPR. This ratio is not entirely appropriate as excess deaths include deaths from young children born during the PPR. Including these PPR births in the denominator, we find that the number of excess deaths represents 21·0% of the population at risk (median value, Table 2, 90% CI: 14.7% to 27.6%). Another way to capture the dramatic increase in mortality during the PPR is to consider life expectancy at birth. Trending upward until the late 1960s, life expectancy at birth for both males and females probably edged over 50 years briefly before falling back slightly below that threshold during the civil war. By contrast, we estimate life expectancy at birth during the PPR to have dropped to as low as 16.7 years for females (median, 90% CI: 10.3 to 28.5 years) and 8.1 years only for males (median, 90% CI: 5.6 to 12.1 years). Finally, we also estimate a concomitant decline in total fertility down to 4.21 children per woman, corresponding to a PPR deficit of 358,000 births (median, 90% CI: 251,000 to 481,000 births).

Discussion

A review of extant PPR-mortality estimates suggests that the most plausible among these estimates range from 720,000 to 1.1 million violent deaths and from 1.0 to 2.1 million excess deaths. However, each of these figures is a point estimate with some unknown uncertainty. As shown in Figure 1, the levels of uncertainty regarding demographic trends make the production of point estimates relatively uninformative. This does not only apply to estimations from demographic evidence, but also to estimations based on samples of survivors' testimonies which require an estimate of the 1975 total population size. In relative terms, estimations based primarily on verbal evidence cannot be more precise than the 1975 population size estimate.

To quantify our knowledge of PPR mortality and fertility, we use the demographic relationships embodied in the CCMPC and our extant knowledge of other demographic parameters between 1962 and 1998, which was incorporated in the input parameter distributions. The distributions represent potential measurement errors and also expert disagreement about the reliability of extant data. Obviously, a different representation of the current consensus about those would yield different frequency distributions for PPR mortality and fertility indicators. Not to overstate this consensus, we defined "generous" distributions, but in fine, an element of subjectivity is not entirely avoidable. Analyses of the output distributions' sensitivity to changes in the parameter (input) distributions did not indicate a particularly influential parameter.

Another issue is whether the model implementation then appropriately derives the PPR mortality and fertility distributions from the input distributions. Ideally, the joint distribution of all the mortality, fertility and migration parameters involved in this reconstruction would have been specified. Unfortunately, demographic theory is relatively poor with respect to the inter-relationships between demographic variables (Cleland 2001) and an assumption of independence between the fertility, mortality and migration parameters seems more easily defensible than any alternate specification. Whether this approach understates or overstates the level of uncertainty depends on the direction of the correlations (here assumed to be zero) between the demographic components. Recent work on the demographic transition, population momentum and homeostasis indicates possible correlations between mortality, fertility and migration (Heuveline 1999; Blue and Espenshade 2011; Billari and Dalla-Zuana 2013) in a direction that reduce overall uncertainty. On period lengths such as those used here (13 years for the forward projection and 18 years for the backward projection), however, these correlations are likely to have but a very minor impact on population dynamics.

As for the underlying demographic model itself, the CCPMC makes only a few, reasonable assumptions and none that should have much impact on the output frequency distributions. More caution is in order regarding the decomposition of causes-of-death based on mortality models. The extraction of non-violent deaths from excess deaths depends on the ability of a high-mortality model to capture the age pattern of these non-violent deaths. To compensate for the possibility that some non-violent deaths could remain among the residual deaths "unexplained" by this model, the level of non-violent mortality was set to the highest plausible level, namely, the level that accounts for all excess deaths in one of the female age group. This unrealistic working assumption arguably entails some over-compensation and the resulting distribution of violent female deaths (with only 11,000 as a lower-bound of our 95% CI) indicates a possible underestimation of violent mortality.

The main conclusion remains that the uncertainty surrounding PPR death toll estimates is substantial. While the current range of death-toll estimates can be narrowed to 1.0 to 2.1 million excess deaths after ruling out a couple of less methodologically sound estimates, the 95% CI derived here is about 50% wider, stretching from 1.201 to 2.757 million excess deaths. The 95%

CI is 728,000 to 1.455 million for violent deaths alone, nearly twice as wide as the range of extant estimates. Wide intervals are to be expected in such contexts: a survey of 1955-2002 war mortality in 13 countries yielded a 95% CI ranging from 3.0 to 8.7-million deaths (Obermeyer et al. 2008).

Moreover, this assessment provides a better basis for comparing the extant estimates than a review of the published information on their approach and data alone. This review suggested that recent estimates tended to be higher than earlier estimates, but could not readily determine whether recent ones were actually more reliable. For those derived from demographic evidence, the results shed light on the trade-off between benefitting from recent, more reliable data and requiring a longer reconstruction period that must guide the choice of a post-PPR data point to anchor the reconstruction. The 95% CI derived here for the total population size at the end of the PPR (6.10 to 6.64 million) compares with the 6.40 million figure from the first governmental estimate in 1980 (Banister and Johnson 1993). Another post-PPR data point that has been used to anchor the backward reconstruction is the number of people aged 20 and over documented in the United Nations roster at the end of 1992 (Heuveline 1998). Given the electoral eligibility criteria that excluded some of the de facto population, the inability to conduct registration in some still disputed areas, and the 96% estimated registration coverage elsewhere (United Nations 1995), the 95% CI (4.69 to 4.80 million) also seems compatible with the 4.28 million aged 20 and over appearing in the electoral rosters. The spreads of the CI even seem narrower than what could reasonably be derived from a careful review of these data. This implies that choosing the 1998 Census data over earlier data sources to anchor the backward reconstruction eventually yields narrower CI on demographic indicators of the PPR period as well, and in this sense, later estimates seem superior to earlier ones. The latest, deterministic reconstruction (Lognard 2012) even uses the 2008 census. A replication in a stochastic framework would be required to determine whether a 10-year backward projection from the 2008 census data would similarly yield less uncertainty about the 1998 population than the 1998 census itself, but this appears unlikely.

Finally, a striking finding is that even wide as the 95% CI derived here may appear, they only cover four of the seven extent estimates of violent mortality and five of the nine extant estimates of the death toll. The lowest three of the seven extant estimates of violent mortality appear most unlikely, all three being even below the lowest value derived in our 10,000 simulations (377,000). As regards the death toll, the reservations expressed on methodological grounds in the review of the lowest (741,000 deaths) and the highest (3.315 million deaths) estimates can now be substantiated. The distribution obtained here indicates only a .58% likelihood that excess deaths were over three million and a .52% likelihood that they were under one million. We also find that two of the remaining seven seemingly plausible estimates in the one to two million death interval fall outside the 95% CI too. Overall, the often–cited interval of one to two million excess deaths only has a 60.19% likelihood. The narrower 1.5 to 2.25 million interval actually has a higher likelihood (68·93%) of including the true value.

In spite of the uncertainty regarding the exact figures, this assessment leaves little doubt that the impact of the PPR on demographic trends was massive. Even considering only violent mortality, the lower bound of the 95% CI indicates that the actual value has a 97.5% likelihood of being at least 728,000 violent deaths. Providing a more comprehensive assessment of the PPR consequences, actual values have again a 97.5% likelihood of being higher than our lower bound estimates of a 1.20-million death toll and a 232,000 birth deficit respectively. This minimum estimate of the death toll would still represent 13.4% of the population at risk and equate 323,000 deaths annually, almost as many (378,000) as a median estimate of global war deaths between 1985 and 1994 (Obermeyer et al. 2008).

Acknowledgments

The author wishes to thank Mark Handcock, Adrian Raftery, Rob Mare, Ron Brookmeyer, Patrick Gerland, Matthew Salganik, and Seth Sanders for their insightful suggestions and Andrew Hicks for implementing model simulations in Stata (release 12). This research made use of the facilities and resources at the California Center for Population Research, UCLA, which is supported by infrastructure grant R24HD041022 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development. The author is solely responsible for the content of this article.

Appendix I: A brief review of extant estimates

Estimates primarily based on verbal evidence

Arguably the first attempt to scale verbal evidence up to assess the impact at the national level, the Central Intelligence Agency (CIA 1980, thereafter) report does not offer a figure for the death toll of the PPR, but rather indicates that population declined by 1.2 to 1.8 million due to "the savagery of that regime." As the decline is the product of higher mortality, lower fertility and induced migration, a number of excess deaths is not readily derived from these figures. Reviewing the CIA report and with additional data on food production, however, Ea (1981) estimates the number of excess deaths between 1.0 and 1.2 million. Neither of them reports on the nature of the data or sample size.

Heder (1997) reaches the estimate of 1.5 million excess deaths based on about 1,000 testimonies collected in Cambodia and in refugee camps at the Thai border in 1980–81. Combining about 400 family interviews also conducted at the border, with about 100 similar interviews collected earlier in France, Kiernan (1996) arrives at 1.671 million. Sliwinski (1995) best documents his verbal record collection in France (60), Thailand (600) and Cambodia (600), which took place a full decade later (1989–91), when access to some of the provinces beyond the capital city had improved substantially, albeit partially still. With two different sets of assumptions, he arrives at two estimates for the number of excess deaths during the PPR, 1.843 and 1.869 million.

In the meantime, the post-PPR government had put forth the official figure of 3.315 million excess deaths (Gordon 2007). Rather than proceeding from a sample, however, the government arrived at the number of excess deaths by tallying up testimonies collected throughout the country in 1982–83 (known as the "Renakse Petitions").

Testimonies have also been used to estimate directly the proportion of deaths that were likely violent deaths. However, Sliwinski's (1995) data show as much as 13% of the PPR-era losses to be "disappeared" people or people whose cause of death was unknown to the reporting person. To the 36% of excess mortality reported as due to violence, one must then add a share of the "disappeared" people or those dead of unknown causes whom might have eventually suffered a violent death. The tentative re-allocation would plausibly yield between 720,000 violent deaths —using his lower estimate of excess deaths and increasing the violent/excess mortality ratio from 36% to 39% —and 840,000 violent deaths—using his higher estimate of excess deaths and increasing the ratio to 45%. Other estimates based on verbal evidence are not as explicitly documented. Heder's (1997) comes close to Sliwinski's, as his reckoning that violent deaths accounted for about 50% of excess mortality would yield 750,000 violent deaths. Much lower is the CIA's (1980) early estimate of 50,000 to 100,000 executions and that of Ea (1981, p.211), who states: "I believe that any estimate exceeding 120,000 deaths by violence, including deaths attributed to the conditions of war with Vietnam, is exaggerated—indeed, totally unrealistic."

Estimates primarily based on demographic evidence

Vickery (1984) used a governmental post-PPR population size estimate to derive the number of excess deaths as a residual term, that is, the difference between a counterfactual and the actual population change. The estimate he used, 6.8 million in 1981, was apparently a projection from the population count the government arrived at in 1980 by tallying up numbers at the local administrative level. The figure is higher than the population size at the end of the PPR estimated in the CIA (1980) report (between 5.4 and 6.2 million), and by Ea (1981) (6.510 million). The high figure was initially suspected of being inflated by local officials in order to increase their share of resources and aid, which in a still very centralized post-PPR Cambodia, were mostly distributed top down. Comparisons with population estimates for later dates do not suggest a massive inflation (including in the present reconstruction, as shown in Fig. 1), and Banister and Johnson (1993) argue that the incentives to exaggerate numbers locally might have been balanced by local registration missing people during what was still a period of relocation, search for relatives and high overall mobility. In any event, Vickery compared the 1981 estimate with his projection of the 1962 population forward to the onset of the PPR (7.64 million). The latter is lower than most other figures for that time, and indeed near the lower bound of the 95% confidence interval derived here (7.62 million, Fig. 1). A high population-size figure for 1981 and a low one for 1975 yields a small residual number, and indeed his estimate that there were "740,800 deaths in excess of normal and due to the special conditions" of the PPR is the lowest to date. He adds that of those excess deaths, a "subjective assessment of survivor's accounts indicates," there are "300,000 to be attributed to executions."

Nearly a decade later but still using official government figures, Banister and Johnson (1993) carry on a full demographic reconstruction by age and gender, yielding an estimate of 1.05 million excess deaths. Without new demographic data, Heuveline (1998) used the 1992 United-Nations (1995) rosters of eligible voters aged 21 and over. (Eligibility started at age 18, but in the absence of proper identification, the number of 18 to 20 year-olds could have been inflated by people under 18 years old having self-reported as older in order to vote.) The age limit implied that his reconstruction only included the birth cohorts before 1970, among which he estimated 2.52 million excess deaths between 1970 and 1979. Excluding excess deaths during the civil war of 1970-75 and the famine believed to have occurred in 1979, and adding those of the post-1970 birth cohorts, he suggests "1.5 to 2.0 million [excess deaths] in 1975-78 alone." (1998, p.60) After the release of the 1998 Census data, Neupert and Prum (2005) eventually estimate 1.953 million excess deaths for the decade, roughly 600,000 below Heuveline's (1998) median estimate for the same period. They do not provide an estimate for 1975–79 alone, but had they done so plausible figures for deaths corresponding to 1970–75 and to 1979 suggest that their estimate would be higher still than Banister and Johnson's. Finally, Lognard (2012) uses both the 1998 and 2008 censuses, taking advantage of them, in particular, to correct the reported age structure of the population. Just above two million, hers is the highest single-point estimate to date.

Of these, only Heuveline (1998) attempts to use the reconstruction to isolate the subset of excess deaths due to violence, by leveraging differences in cause-specific age patterns of mortality. Though the approach also provides a low estimate of 600,000 and a high estimate of 2.0 million, his median estimate (1.1 million) is higher than those produced from verbal evidence.

Estimates primarily based on physical evidence

In April 1994, the U.S. Congress passed the "Cambodian Genocide Justice Act," intended "to support efforts to bring to justice members of the Khmer Rouge for their crimes against humanity committed in Cambodia between April 17, 1975 and January 7, 1979." After being awarded funds by the U.S. State Department to document the PPR crimes, Yale University established the Documentation Center of Cambodia (DC-Cam) in Phnom Penh—since 1997, a permanent, autonomous research institute. The DC-Cam has cataloged printed materials on and photographs from the PPR, and attempted to construct an exhaustive inventory of prisons and mass grave pits. Combining global satellite position mapping and on-the-ground interviews with local informants, DC-Cam had uncovered 20,442 mass graves by 2004. At most sites, DC-Cam fieldworkers were able identify local witness that claimed that the graves contained victims brought there by the

PPR security forces and killed in a nearby prison or on site. The tally corresponding to these graves stands at 1,112,829 bodies (Etcheson 2005). The number may continue to increase as the inventory of mass graves becomes more complete, but whether the corresponding death toll only includes violent deaths or also includes deaths from other causes remains a contentious issue. Forensic analyses were anticipated to determine the cause of death, but due to budget limitations, they could be used at no more than a small fraction of the graves.

Appendix II: Parameters of the demographic reconstruction

The distributions of the 47 parameters in the demographic reconstruction are summarized in Appendix Table A1. Further details on the sources reviewed and on setting these parametric distributions are available as supplementary materials.

References

Banister, Judith and E. Paige Johnson. 1993. After the nightmare: The population of Cambodia, in Benedict Kiernan (ed), *Genocide and Democracy in Cambodia: The Khmer Rouge, the United Nations and the International Community*. New Haven, CT: Yale University Southeast Asia Studies, pp. 65–140 .

Billari, Francesco C. and Gianpiero Dalla-Zuana. 2013. Cohort replacement and homeostasis in world population, 1950–2100, *Population and Development Review* 39(4): 563–85.

Blue, Laura and Thomas Espenshade. 2011. Population momentum across the demographic transition, *Population and Development Review* 37(4): 721–47.

Central Intelligence Agency. 1980. *Kampuchea: A Demographic Catastrophe*. Research Paper GC80-10019U. Washington, DC: National Foreign Assessment Center.

Chandler, David. 1991. *The Tragedy of Cambodian History: Politics, War and Revolution since 1945*. New Haven, CT: Yale University.

Cleland, John. 2001. The effects of improved survival on fertility: A reassessment, *Population and Development Review* 27(Supplement: Global Fertility Transition): 60–92.

Ea, Meng Try. 1981. Kampuchea: A country adrift, *Population and Development Review* 7(2): 209–228.

Etcheson, Craig. 2005. *After the Killing Fields: Lessons from the Cambodian Genocide*. Westport, CT: Praeger.

Gakidou, Emmanuella and Gary King. 2003. Death by survey: Estimating adult mortality without selection bias from sibling surviving data, *Demography* 43(3): 569–585.

Gordon, Amy. 2007. The Renakse petitions: Background and suggestions for future use, manuscript, Documentation Center of Cambodia (Phnom Penh, Cambodia).

Hagan, John and Alberto Palloni. 2005. Death in Darfur, *Science* 313: 1578–1579.

Handcock, Mark S. 1989. *Latin Hypercube Sampling to Improve the Efficiency of Monte Carlo Simulations: Application to a CSET Logic Circuit with ASTAP*. IBM Research Report, RC 14546. Yorktown Heights, NY: Thomas J. Watson Research Center.

Heder, Stephen. 1997. Personal communication.

Heuveline, Patrick. 1998. 'Between one and three million': Towards the demographic reconstruction of a decade of Cambodian history, *Population Studies* 52(1): 49–65.

Heuveline, Patrick. 1999. The global and regional impact of mortality and fertility transitions, 1950–2000, *Population and Development Review* 25(4): 681–702.

Heuveline, Patrick. 2001. Approaches to measuring genocide: Excess mortality during the Khmer Rouge period, in Daniel Chirot and Martin Seligman (eds), *Ethnopolitical Warfare: Causes, Consequences, and Possible Solutions*. Washington, DC: American Psychological Association, pp. 93–108.

Heuveline, Patrick and Samuel J. Clark. 2011. Model schedules of mortality, in Richard G. Rogers and Eileen M. Crimmins (eds), *International Handbook of Adult Mortality*. New York: Springer, pp.511–532.

Heuveline, Patrick and Bunnak Poch. 2007. The phoenix population: Demographic crisis and rebound in Cambodia, *Demography* 44(2): 405–426.

Iman, Ronald L. and Jon C. Helton. 1988. An investigation of uncertainty and sensitivity analysis techniques for computer models, *Risk Analysis* 8(1): 71–90.

Kiernan, Benedict. 1996. *The Pol Pot Regime*. New Haven, CT: Yale University.

Lee, R. D. 1985. Inverse projection and back projection: Comparative results and sensitivity tests for England, *Population Studies* 39(2): 233–248.

Lognard, Marie-Odile. 2012. "Conséquences démographiques de la période khmer rouge," [Demographic consequences of the Khmer Rouge period] paper presented at the Chaire Quételet, Impact démographique des conflits. Last accessed on 12/12/2013 at http://www.uclouvain.be/cps/ucl/doc/demo/documents/Lognard.pdf

Neupert, Ricardo F. and Virak Prum. 2005. Cambodia: Reconstructing the demographic stab of the past and forecasting the demographic scar of the future, *European Journal of Population* 21(2/3): 217–243.

Obermeyer, Ziad, Christopher J. L. Murray, and Emmanuella Gakidou. 2008. Fifty years of violent war deaths from Vietnam to Bosnia: Analysis of data from World Health Survey Programme, *British Medical Journal* 336(7659): 1482–1486.

Preston, Samuel H., Antonio McDaniel, and Carlos Grushka. 1993. "New model life tables for high-mortality populations," Historical Methods 26(4): 149-59

Preston, Samuel H., Patrick Heuveline, and Michel Guillot. 2001. *Demography: Measuring and Modeling Population Processes*. Malden, MA: Blackwell.

Seltzer, William. 2008. "Exploring the limits of demography: The use of "excess mortality" to estimate genocide mortality," manuscript, Fordham University.

Shryock, Henry S. and Jacob S. Siegel. 1973. *The Methods and Materials of Demography*. Washington, DC: U.S. Department of Commerce.

Sliwinski, Marek. 1995. *Le génocide Khmer Rouge: Une analyse démographique*. [The Khmer Rouge Genocide: A Demographic Analysis.] Paris : L'Harmattan.

Trussell, John and German Rodriguez. 1990. A note on the sisterhood estimator of maternal mortality, *Studies in Family Planning* 21(6): 344–346.

United Nations. 1951. Convention on the prevention and punishment of the crime of Genocide. Adopted by the General Assembly of the United Nations on 9 December 1948, *Treaty Series* 78-I-1021: 277-322.

United Nations. 1995. *The United Nations and Cambodia, 1991–1995,* The Blue Book Series. New York, NY: United Nations.

Vickery, Michael. 1984. *Cambodia 1975–1982*. Boston, MA: South End Press.

Weitz, Eric D. 2003. *A Century of Genocide*. Princeton, NJ: Princeton University.

Wheldon, Mark C., Adrian E. Raftery, Samuel J. Clark, and Patrick Gerland. 2013. Reconstructing past populations with uncertainty from fragmentary data, *Journal of the American Statistical Association* 108(501): 96-110.

Table 1: Extant estimates of mortality during the Pol Pot regime, by main evidence type and methodology

Main evidence type	Methodology	Author	Estimates: Violent Deaths	Estimates: Excess Deaths	Estimates: Other
Verbal (survivors' Testimonies on kin survival/ death)	Sample probabilities of death applied to sub-population totals	CIA	50 - 100 (executions)	n/a	Population decline, 1975-79: 1,200 - 1,800
		Ea	100 - 120	1,000 - 1,200	
		Kiernan	n/a	1,671	
		Heder	750	1,500	
		Sliwinski	720 - 840	1,843 - 1,869	
	Sum of death reports	Cambodian Government	n/a	3,315	

	Total accounting	Vickery	300	741	
Demographic (census and national surveys)	Demographic reconstruction by age and sex	Banister & Johnson	n/a	1,050	Excess deaths, 1970-79: 1,300
		Heuveline	1,100	1,500 - 2,000	Excess deaths among cohorts born before 1970, 1970-79: 2,520
		Neupert & Prum	n/a	n/a	Excess deaths, 1970-79: 1,953
		Lognard	n/a	2,095	
Physical (bone remains)	Inventory of mass graves	Documentation Center of Cambodia	1,113	n/a	

Note: See Appendix I for further details about extant estimates

Table 2: Mortality and fertility indicators, Pol Pot regime (April 17th, 1975 to January 7th, 1979)

	Median	Average	Std Dev.	90% CI		95% CI	
Violent deaths (in thousands)							
Total males	884	884	93	729	1,037	703	1,066
Total females	205	204	99	41	368	11	398
Total	1,090	1,088	186	783	1,398	728	1,455
Excess deaths (in thousands)							
Total males	1,270	1,275	186	978	1,610	919	1,696
Total females	631	633	193	326	951	275	1,058
Total	1,902	1,908	374	1,317	2,548	1,201	2,757
Excess deaths relative to population at risk (in percent)	21.0	21.0	3.9	14.7	27.6	13.4	29.8
Life expectancy at birth (in years)							
Male	8.1	8.4	2.1	5.6	12.1	5.2	13.2
Female	16.7	17.8	5.8	10.3	28.5	9.3	31.6
Total fertility rate (children per woman)	4.21	4.22	0.48	3.46	5.03	3.32	5.19
Birth deficit (in thousands)	358	360	69	251	481	232	504

Appendix Table A1: Parameters' description and value distribution

Parameter Description	**Value Distribution**
Forward Projection (1962-75)	
Initial total population size, 4/17/1962: Undercount in the 1962-Census official total (%)	N(3.2,1)
Initial sex and age structure, 4/17/1962: Weighting function between the official (when weight=-1) and Siampos' (1970) 1962-Census adjustment (when weight=+1)	N(0,1)
Mortality age pattern, 1962-67: Weighting function between UNSA (when weight=0) and CDW (when weight=1) mortality patterns	U(0,1)
Initial mortality level reference date: Shift in reference date for the first male and female life expectancy estimates (years)	N(0,2)
Male mortality level, 1962-67: Average annual gain in male life expectancy at birth (years)	N(.5,.1)
Female mortality level, 1962-67: Female v. male difference in average annual gain in life expectancy at birth (years)	N(.1,.1)
Mortality age pattern, 1967-70: Convergence toward CDW between 1962-67 & 1967-70 (with no change from 1962-67 when weight=0 and CDW when weight=1)	U(0,1)
Male mortality level, 1967-70: 1962-67 v. 1967-70 difference in average annual gain in male life expectancy at birth (years)	N(.1,.05)
Female mortality level, 1967-70: Male v. female difference in 1962-67 v. 1967-70 difference in average annual gain in life expectancy at birth (years)	N(.05,.05)
Fertility level, 1962-67: Correction factor to the 1962-Census TFR estimate by Siampos	N(-.2,.2)
Fertility level, 1967-70: Annual change in TFR between 1962-67 & 1967-70	N(-.1,.1)
International migration, 1962-70: Emigrants during the first 8 months of 1970 (thousands)	U(150,300)
Counterfactual male mortality level, 1970-75: Ratio of average annual life expectancy change in baseline mortality between 1970-75 & 1967-70 to average annual life expectancy change in total mortality between 1967-70 & 1962-67 (years)	U(-1,1)
Counterfactual female mortality level, 1970-75: Female to male relative ratio of average annual life expectancy change in baseline mortality between 1970-75 & 1967-70 to average annual life expectancy change in total mortality between 1967-70 & 1962-67 (years)	U(0,1)
Excess mortality, 1970-75: Number of excess deaths from bombing, civil-war and pogroms (thousands)	U(50,250)
Domestic fertility level, 1970-75: Relative change in TFR between 1967-70 & 1970-75	U(-.1,0)
Expatriate fertility level, 1970-75: Ratio of change in TFR between 1967-70 & 1970-75 to change between 1962-67 & 1967-70	N(0,1)

International migration, 1970-75: Number of emigrants (thousands)	U(75,175)
Backward Projection	
Initial total population size, 3/2/1998: Undercount in the 1998-Census official total (%)	N(3.2,1)
Initial sex and age structure, 3/2/1998: Inverse of the fraction of each five-year age group shifted to the next, younger five-year age group in the official 1998-Census distribution	N(30,10)
Mortality age pattern, 1993-98: Weighting function between CDN (when weight=0) and CDW (when weight=1) mortality patterns	U(0,1)
Mortality level, 1993-98: Correction to the 1998-Census estimate of life expectancy at birth for both sexes (years)	N(0,1.5)
Gender gap in mortality levels, 1993-98: Female v. male difference in life expectancy at birth (years)	N(4,1)
International migration, 1993-1998: In v. out difference in number of international migrants (thousands)	N(75,25)
International migration, 1993: Number of refugees returning from camps, late 1992 to early 1993 (thousands)	N(400,25)
Mortality age pattern, 1988-93: Convergence toward CDN between 1993-98 & 1988-93 (with no change from 1993-98 when weight=0 and CDN when weight=1)	U(0,1)
Male mortality level, 1988-93: Average annual gain in male life expectancy at birth between 1988-93 & 1993-98 (years)	N(.3,.1)
Female mortality level, 1988-93: Female v. male difference in average annual gain in life expectancy at birth between 1988-93 & 1993-98(years)	N(.1,.1)
International migration, Vietnam, 1979-1993: Number of immigrants (including return migrants) from Vietnam between mid-1979 & March 1993 (thousands)	N(350,75)
International migration, 1988-1993: Proportion of the number of immigrants from Vietnam between mid-1979 & March 1993 occurring in 1988-93	N(.25,.05)
Mortality age pattern, 1983-88: Convergence toward CDN between 1988-93& 1983-88 (with no change from 1988-93 when weight=0 and CDN when weight=1)	U(0,1)
Male mortality level, 1983-88: Average annual gain in male life expectancy at birth between 1983-88 & 1988-93(years)	N(.4,.1)
Female mortality level, 1983-88: Female v. Male difference in average annual gain in life expectancy at birth between 1983-88 & 1988-93 (years)	N(.1,.1)
Immigration, 1983-1988: Proportion of the number of immigrants from Vietnam between mid-1979 & March 1993 occurring in 1983-88	N(.35,.05)
Emigration, 1983-1988: Number of emigrants to Thailand (thousands)	N(300,50)
Counterfactual mortality age pattern, 1979-83: Convergence toward CDN between	U(0,1)

1983-88 & 1979-83 in baseline mortality (with no change from 1983-88 when weight=0 and CDW when weight=1)	
Counterfactual male mortality level, 1979-83: Average annual life expectancy gain in baseline male mortality between 1979-83 & 1983-88 (years)	N(.4,.1)
Counterfactual female mortality level, 1979-83: Female v. male difference in average annual life expectancy gain in baseline mortality between 1979-83 and 1983-88 (years)	N(.1,.1)
Immigration, 1979-1983: Proportion of the number of immigrants from Vietnam between mid-1979 and March 1993 occurring in 1979-83	N(.4,.05)
Emigration, 1979-1983: Number of emigrants to Thailand (thousands)	N(50,15)
Excess mortality, 1979-80: Number of deaths from famine in late 1979 to early 1980 (thousands)	N(150,50)
International migration, 1979: Number of immigrants from Vietnam, first months of 1979 (thousands)	N(100,25)
1975-79 Reconciliation	
Domestic fertility level, 1975-79: Relative change in TFR between 1970-75 & 1975-79	N(.3,.05)
Expatriate fertility level, 1975-79: Ratio of change in TFR between 1970-75 & 1975-79 relative to change in TFR between 1967-70 & 1970-75	N(0,1)
International migration, 1975-78: Number of emigrants (thousands)	U(100,300)
International migration, 1978-79: Number of emigrants, ca. January 1979 (thousands)	N(150,25)
Excess mortality timing, 1975-79: fraction of the difference in 1979 population by sex and age between the counterfactual and the backward projections in 1979 that are attributed to excess deaths in the second half of the PPR	U(.5,1)

Note: UNSA and CDW model patterns refer to the United Nations' (1982) and Coale and Demeny's (1983) model life tables.

Figure Titles

Figure 1: Cambodia's Total Population Size, 1962 to 1998, Median and 95% Confidence Interval (in millions)

Figure 2: Pol Pot regime's death toll, estimates' density distribution (number of excess deaths, in thousands)