eAppendix

Correction of gestational age estimates

A common way to clean gestational-age data is to exclude or reclassify babies whose birth weight is incompatible with the ostensible gestational age. We used birth weight *z*-scores to make this judgment. Z-scores require gestational-age-specific estimates of birth weight standard deviations, which are rarely published. Kramer et al[1] have provided corrected week-specific birth weight standard deviations based on singleton births from Canada, where the great majority of pregnant women undergo an ultrasound early in pregnancy[1]. We made use of these external standards, applying them to twin and triplet births as well as singleton, while using the median based on our own data to estimate the week-specific mean birth weight for singletons, twins, and triplets separately. Of the 31,247,865 babies eligible for analysis (figure 3 in main paper), 30,328,934 (97.1%) were singletons and 918,931 were twins or triplets.

Most babies had both LMP and "clinical" estimates (presumably based on ultrasound in most cases). In the 4.9% of instances where the LMP was derived from the clinical estimate, we retained only the latter. We retained babies with imputed day of LMP (7%) or imputed birth weight (0.3%).

We focused on the window of gestation between 24 and 44 weeks. If both gestational-age measures were outside this window, babies were excluded. If only one of the two measures was outside the window, we set that value to missing. These babies were considered as having only one "valid" estimate of gestation (0.9%).

We calculated birth weight z scores using the standard deviations (SD) estimated by Kramer et al[1], after calculating a pooled value independent of baby's sex (and using the SD of 43 weeks for both 43 and 44 weeks). We performed all analyses using 250-g categories of birth weight. At some gestations, mean birth weight is heavily affected by the presence of misclassified babies in the right tail of the distribution. For this reason, we used the median at each gestation as a proxy for the mean. The median was calculated separately for singletons, twins, and triplets using the clinical estimate of gestation at preterm weeks, and LMP for term babies. (For triplets, estimates were based on LMP from week 36).

Construction of an algorithm for cleaning gestational-age data is complex. In order to make our decisions as clear as possible, we provide a diagram (Appendix figure 1) to describe the steps through which we retained or excluded babies (see legend). To evaluate whether the z-score was appropriate for any given gestation, we set different criteria for term and preterm babies. The common criterion for all babies with an estimate of 37 weeks or higher was that birth weight *z* score be between -5 and +5 SD. We used two criteria for preterm babies: the looser criterion, requiring the *z* score to be between -4 and +3 (marked with * in appendix figure 1), was applied to babies with estimates agreeing within two weeks of each other, and also when checking clinical gestation in babies whose estimates were more than 2 weeks apart. The stricter criterion (marked with ** in appendix figure 1), requiring z to be between -3 and +2, was applied when checking the LMP of babies with estimates of gestation that disagreed by more than 2 weeks (if clinical gestation had not fulfilled the looser criterion) and to those with only one estimate of gestation. We used asymmetric z-score criteria for birth weight mainly because the right tail of heavy babies is the dominant problem in the misclassification of preterm gestational weeks.

We excluded a total of 66,561 babies (0.2%) whose birth weights were incompatible with the available information on gestational age (0.2% among singletons, 0.3% among twins and triplets). Appendix Table 1 shows the proportion of LMP values that were accepted, rejected, or reassigned (downward or upward) among among singletons with an LMP-based estimate. Estimates between 28 and 35 weeks were those most likely to be discarded or reassigned upwards. Estimates from 42 to 44 weeks were those most likely to be reassigned downward.

The cleaning process markedly reduced the estimated proportion of singleton preterm births: among the retained babies, 8.3% were preterm (among those with a gestational age between 24 and 44 weeks) using the “cleaned” gestational age compared with 10.0% using the LMP-based estimate.

In twins and triplets, the preterm birth rates were increased slightly (57% in cleaned data compared with 55% using LMP, and 95 % compared with 92%). As expected, mortality was higher among the excluded babies.[2,3]

Such data-cleaning markedly simplifies the birth weight distributions and gestational-age-specific mortality curves. Appendix figure 2 shows the distribution of birth weight at 30 and 32 weeks, before and after cleaning, for singletons, twins, and triplets. The “before” curve is based on LMP, including also the values estimated on the basis of clinical gestation when LMP is missing (as these data would be customarily used). The “after” curve is based on whichever estimate (LMP or clinical) was preferred during the data-cleaning process, and excluding those whose birth weights failed to meet the z-score criteria. In singletons, the “after” curve appears somewhat truncated, but the right heavy tails have largely disappeared.

The uncorrected birth weight distributions of twins and triplets were less affected by a heavy right tail than that of singletons at virtually all preterm gestations. This suggests that LMP-based estimates are generally more reliable for twins and triplets than for singletons. This may reflect the fact that many of these babies are the result of infertility treatment. \

We assessed twins and triplets using the standard deviation for birth weight that had been derived from singletons (Kramer et al did not report such estimates for twins or triplets[1]). The true standard deviation for multiples may be different from that of singletons. In our data, when using the clinical estimate of gestation, multiples at all preterm weeks had a smaller estimated

SD than singletons. (This could reflect more errors in clinical gestation among singletons than among multiples, although such errors were not as evident with clinical estimates as with LMP.) If our applied SD overestimated the true SD for multiples, then the algorithm would have resulted in retaining more extreme birth weight values at each gestation. Even so, the LMP birth weight distributions of twins and triplets at preterm weeks were less affected by the right tail problem seen in singletons.

Appendix figure 3 shows the weight-specific mortality curves at 30 and 32 weeks before and after data-cleaning. These figures suggest that most of the excluded births truly had a higher gestation.

After correction, gestational-age-specific neonatal mortality rates were between 92% and 140% of those based on LMP (data not shown). The largest differences in mortality (between +25% and +43%) were seen between 28 and 34 weeks.

Appendix figure 4 shows the gestational-age-specific mortality for singletons before and after cleaning. This shows a higher neonatal mortality at early gestational ages with the corrected data, as expected given the pattern of errors.

In sum, we believe this algorithm for cleaning gestational-age data strikes a reasonable compromise between excluding and reassigning a minimum number of babies, and substantially improving the quality of the remaining data. This method would in principle be useful in any setting where two measures of gestational age are available, and where there are reasonable external standards for gestational-age-specific birth weight SDs.

References

1. Kramer MS, Platt RW, Wen SW, Joseph KS, Allen A, Abrahamowicz M, Blondel B, Breart G. A new and improved population-based Canadian reference for birth weight for gestational age. *Pediatrics* 2001;108(2):E35.
2. Lynch CD, Zhang J. The research implications of the selection of a gestational age estimation method. *Paediatr Perinat Epidemiol* 2007;21 Suppl 2:86-96.
3. Parker JD, Schoendorf KC. Implications of cleaning gestational age data. *Paediatr Perinat Epidemiol* 2002;16(2):181-7.

eTable 1. Proportion of singleton babies whose LMP estimate of gestational age at birth was reassigned or excluded, by reported LMP week of gestation.

Week	N	% Unchanged	% Reassigned downward	% Reassigned upward	% Discarded
24	20,616	79.50	0.00	16.91	3.59
25	24,606	79.45	0.00	17.27	3.28
26	29,362	77.70	0.00	18.69	3.61
27	32,241	74.16	2.23	19.61	4.00
28	45,218	60.97	2.86	30.90	5.27
29	55,831	57.59	3.21	34.43	4.77
30	77,594	53.46	3.10	37.47	5.96
31	101,108	54.30	2.65	37.76	5.29
32	139,714	56.56	1.48	37.36	4.59
33	209,649	58.18	1.43	36.87	3.52
34	368,121	61.22	1.27	34.85	2.66
35	623,477	69.35	1.20	28.65	0.81
36	1,115,674	79.95	0.91	18.69	0.45
37	2,249,351	89.01	0.75	10.21	0.03
38	4,681,997	97.58	0.68	1.72	0.02
39	7,045,008	98.65	1.03	0.30	0.02
40	6,293,019	97.94	2.01	0.03	0.02
41	3,409,058	93.53	6.43	0.01	0.02
42	1,163,621	74.81	25.18	0.00	0.01
43	575,277	34.89	65.10	0.00	0.01
44	309,144	18.98	81.01	0.00	0.01
Total	28,569,686	90.98	4.97	3.85	0.20

The table does not include 1,759,248 (6.2%) singletons for whom only clinical gestation was available (due to missing LMP, LMP imputed from clinical gestation, or LMP <24 weeks or >44 weeks). Babies with a missing estimate of clinical gestation (N=10,637) are included in the table, but could only be either accepted or discarded, not reassigned.

eFigure 1. Steps in attribution of gestational age.

Legend:

*For term babies: gestational-age-specific *z* between -5 and +5.

For preterm babies: gestational-age-specific *z* between -4 and +3.

**For term babies: same as *.

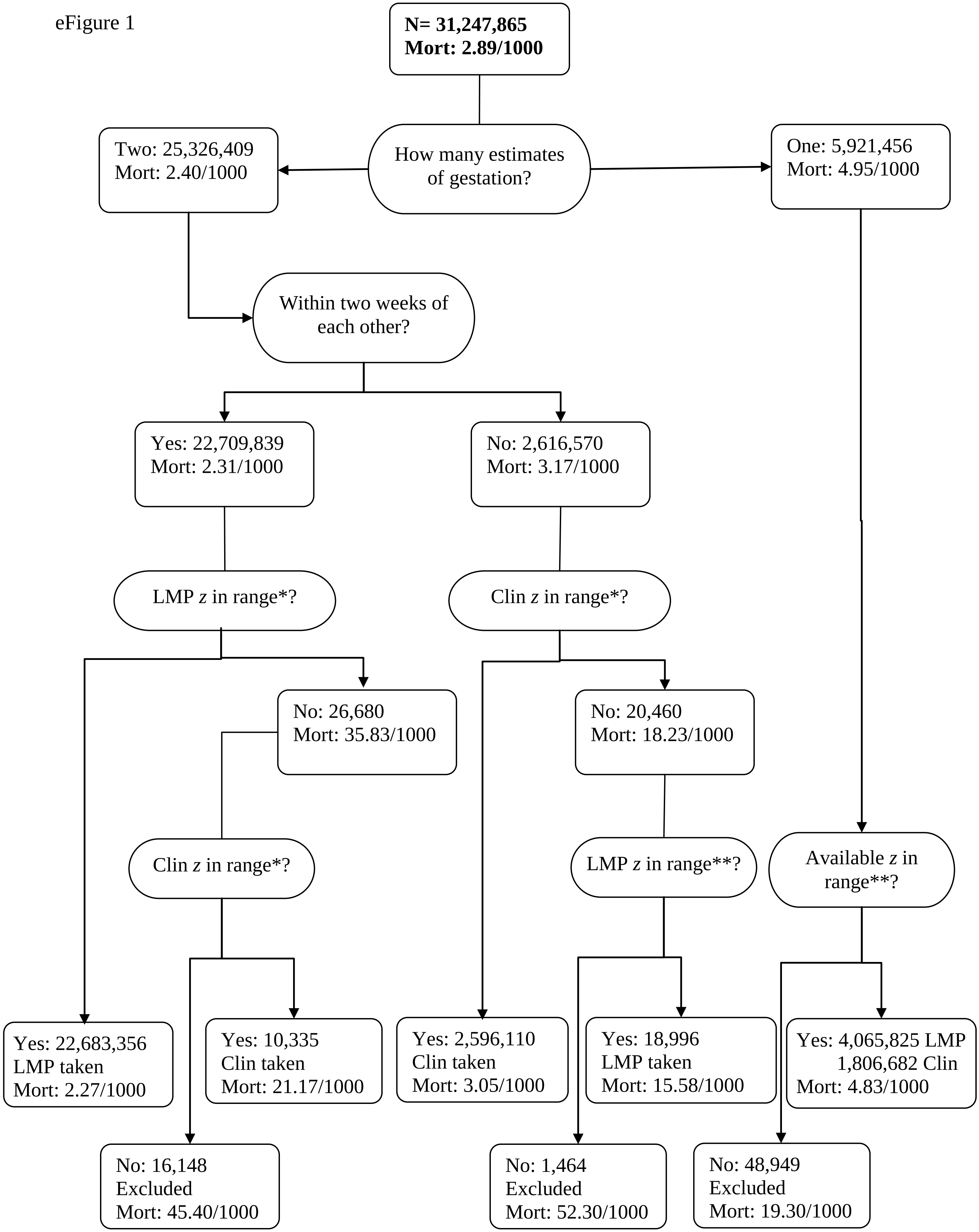
For preterm babies: gestational-age-specific *z* between -3 and +2.

Among babies with two estimates of gestational age, LMP was accepted if the two estimates agreed within 2 weeks and z fulfilled the looser criterion * (N=22,683,159). If *z* was not in the acceptable range, the clinical estimate of gestation was examined against the *z* based on clinical gestation and accepted if in the range defined by * (N=10,532); otherwise, the baby was excluded (N=16,148).

For those for whom the two estimates disagreed by >2 weeks, clinical gestation was checked first and accepted if *z* fulfilled criterion * (N=2,596,110). If not, z based on LMP was checked against the LMP estimate (criterion **). (18,996 babies were accepted, and 1,464 excluded).

For babies with only one estimate of gestation we applied the stricter criterion ** to all preterm births (regardless of whether LMP or the clinical estimate was available), and we accepted 4,065,825 babies with LMP and 1,806,682 with clinical gestation, excluding 48,236 babies.

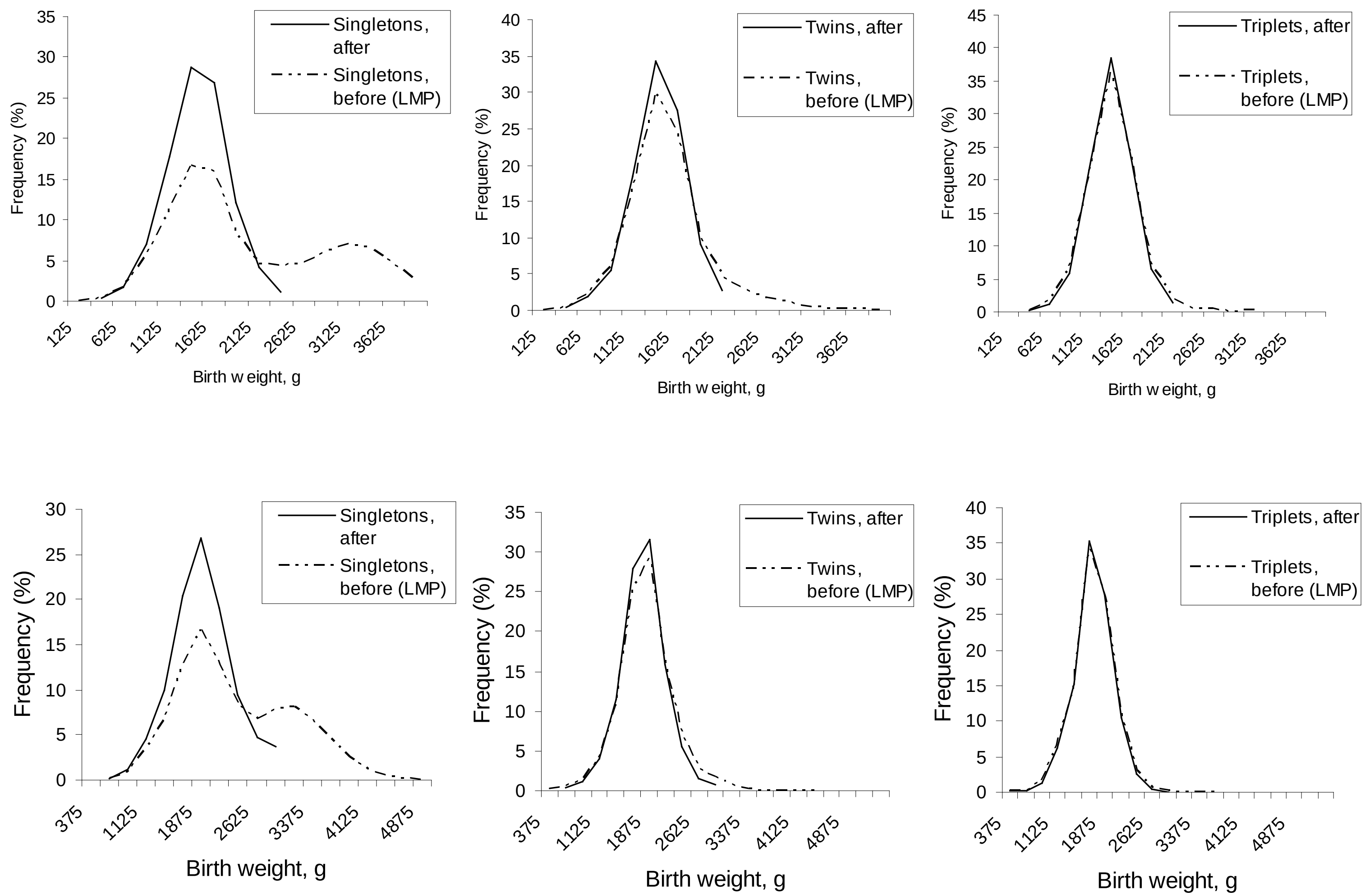
eFigure 1



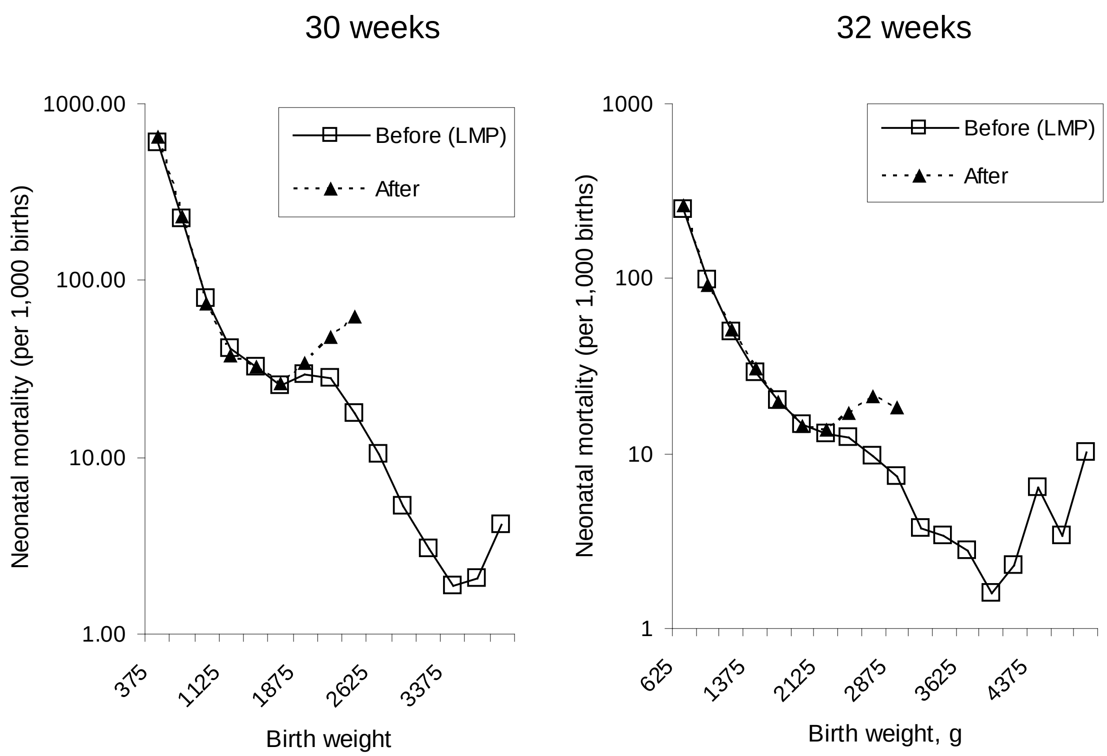
N= 31,247,865
Mort: 2.89/1000
How many estimates of gestation?
Two: 25,326,409
Mort: 2.40/1000
One: 5,921,456
Mort: 4.95/1000
Within two weeks of each other?
Yes: 22,709,839
Mort: 2.31/1000
No: 2,616,570
Mort: 3.17/1000
LMP z in range*?
Clin z in range*?
No: 26,680
Mort: 35.83/1000
No: 20,460
Mort: 18.23/1000
Clin z in range*?
LMP z in range**?
Available z in range**?
Yes: 22,683,356
LMP taken
Mort: 2.27/1000
Yes: 10,335
Clin taken
Mort: 21.17/1000
Yes: 2,596,110
Clin taken
Mort: 3.05/1000
Yes: 18,996
LMP taken
Mort: 15.58/1000
Yes: 4,065,825 LMP
1,806,682 Clin
Mort: 4.83/1000
No: 16,148
Excluded
Mort: 45.40/1000
No: 1,464
Excluded
Mort: 52.30/1000
No: 48,949
Excluded
Mort: 19.30/1000

eFigure 2. Distribution of birth weight for singletons, twins, and triplets before and after cleaning.

Top: Week 30. Bottom: week 32.



eFigure 3. Examples of birth-weight-specific mortality of singletons before and after “cleaning”.



eFigure 4. Gestational-age-specific neonatal mortality of singleton babies before cleaning (based on LMP) and after.