Appendix S1. Search string and systematic searches for national reports

1. perinat*
2. neonat*
3. fet*
4. foet*
5. intrapartum
6. intrauterine
7. intra-uterine
8. 1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7
9. death*
10. dead
11. mortal*
12. demise
13. 9 OR 10 OR 11 OR 12
14. stillb*
15. 8 AND 13
16. 14 OR 15
17. audit*
18. inter-rater*
19. interrater*
20. classif*
21. caus* adj4 (death OR mortal* OR stillb*)
22. 17 OR 18 OR 19 OR 20 OR 21
23. 16 AND 22

Searches were undertaken for national reports and/or statistical data on perinatal mortality from countries across HIC, MIC and LIC, identified from the World Bank Group[1]. Reports were included where in-depth classification of cause of death was reported.

For routine national data, websites of the national statistical office and ministry of health were searched for the following countries: Australia, Canada, Chile, Croatia, France, Ireland, Japan, Kuwait, Lithuania, New Zealand, Poland, Portugal, Qatar, Sweden, United Kingdom, Argentina, Brazil, China, Colombia, Costa Rica, Ecuador, Mexico, Panama, South Africa, Thailand, Ghana, Guatemala, India, Nepal, Zambia, Bangladesh and Ethiopia. For each site, the key phrases: 'stillb*', 'fetal', and 'perinat*' were searched.

For countries with less accessible routine perinatal mortality data, a structured search was conducted within Google with key phrases searched in conjunction with each country. The key phrases included: stillbirth, fetal death, vital statistics; national data; rate/prevalence, statistics.

Additional national report data from Suriname was provided from Dr Hannah Blencowe from The Lancet Stillbirth Epidemiology Investigator Group [2].

Appendix S2. Data collection and definitions

Data collected:

The following data items were extracted for each of the included reports: country; language; year of data collection; setting (hospital (single- or multi-centre) or population based); stillbirth rate; total number of births; number of stillbirths in the cohort and numbers of stillbirths classified; whether termination of pregnancy and multiple pregnancies were included; the reported cause of death (verbatim); the type of data used to assign the cause of death; rates of autopsy and placental pathology; classification system used - name, whether the system used a hierarchical approach, the number of categories, whether the system was aligned with ICD-PM.

Definitions:

Alignment with ICD-PM: Data to assess alignment with ICD-PM was collected as follows: Whether the timing of death (antepartum or intrapartum) and a maternal as well as a fetal condition was identified for each case and whether ICD codes were used for these conditions.

Type of report: Reports were considered population-based if they reported national data or a total cohort of stillbirths within a defined region/district.

Clinical classification system: A clinical classification system was defined as "Any approach to classifying causes of stillbirths described by the authors of included publications as a 'system' or 'approach', and/or that included a clearly delineated list of causes separate from the data"[3].

Hierarchical: A system was considered hierarchical if it required causes to be assigned via consideration of each cause in sequence[3] and partially hierarchical if hierarchy was optional or incompletely defined.

Appendix S3. Checklist for quality assessment

	Yes	No	Unclear
1. Was the sample representative of the target population?			
2. Was the data analysis conducted with sufficient coverage of the identified sample?			
3. Were the study subjects described in detail?			
4. Were objective, standard criteria used for the measurement of the condition?			
5. Was the condition measured reliably?			
a. Adequate investigation of stillbirth?			
b. Adequate data source?			
c. Valid assignment?			

Adapted from the Johanna Briggs Institute Critical Appraisal Checklist for Studies Reporting Prevalence Data[4]

1. Was the sample representative of the target population?

This question relies upon knowledge of the broader characteristics of the population of interest. For this study of causes of stillbirth, knowledge of at least the characteristics, demographics and medical history is needed. The term "target population" should not be taken to infer every individual from everywhere or with similar disease or exposure characteristics. Instead, give consideration to specific population characteristics in the study, including age range, gender, morbidities, medications, and other potentially influential factors. For example, a sample may not be representative of the target population if a certain group has been used (such as stillbirths occurring in hospital, or outside of hospital) and the results then inferred to the target population (i.e. whole population).

Rules:

Answer Yes if: a population based study.

Answer No if: population based study, but there was systematic exclusion that would have meant the cohort is not representative of the target population – for this study this means mainly congenital anomaly

2. Was the data analysis conducted with sufficient coverage of the identified sample?

A large number of dropouts, refusals or "not founds" amongst selected subjects may diminish a study's validity, as can low response rates for survey studies.

- Did the authors describe the reasons for non-response and compare persons in the study to those not in the study, particularly with regards to their socio-demographic characteristics?
- Could the not-responders have led to an underestimate of prevalence of the disease or condition under investigation?
- If reasons for non-response appear to be unrelated to the outcome measured and the characteristics of non-responders are comparable to those in the study, the researchers may be able to justify a more modest response rate.
- Did the means of assessment or measurement negatively affect the response rate (measurement should be easily accessible, conveniently timed for participants, acceptable in length and suitable in content).

Rules: **Answer Yes if:** causes of stillbirth were missing for <20% of the cohort

3. Were the study subjects and setting described in detail?

Certain diseases or conditions vary in prevalence across different geographic regions and populations (e.g. socioeconomic and maternal variables between countries and birth setting). Has the study sample been described in sufficient detail so that other researchers can determine if it is comparable to the population of interest to them?

Rules: **Answer Yes if:** the definition of stillbirth was provided and clear

4. Were objective, standard criteria used for the measurement of the condition?

Here we are looking for measurement or classification bias. Causes of stillbirth can be classified using different types of classification systems or International Statistical Classification of Diseases and Related Health Problems (ICD). The causes assigned should also reflect the underlying cause and not only state conditions (e.g. ICD).

Rules: **Answer is Yes if:**

- ICD was used

  OR

- clinical system which provided good definitions and rules for use (based on Leisher et al[3])

  OR

- the study used an informal list of conditions and included definitions and rules that enable anyone to apply the system to stillbirths

5. Was the condition measured reliably?

Considerable judgment is required to determine the presence of some health outcomes. Having established the objectivity of the outcome measurement instrument, it is important to establish how the measurement was conducted.

Rules: If Yes to all then answer Yes to overall item. Otherwise use the majority of No or Unclear as the answer. If Yes on two items including item 5a, Yes can be assigned to the overall criterion.

5a: Adequate investigation of stillbirth?

Answer Yes if:

- HIC: both autopsy and placenta pathology rates >75%.

  OR

- LMIC Verbal Autopsy was performed in LMIC.

5b: Adequate data source?

Answer Yes if:

- Verbal Autopsy

  OR

- Prospectively collected clinical data for the purposes of classification of causes of death (No if death certificate data only used and/or Vital registration data)

5c: Valid assignment?

Answer Yes if: <50% unexplained and <20% Other unspecified

Overall quality rating algorithm

HIGH QUALITY REPORTS: YES on all criteria.

MEDIUM QUALITY REPORTS: must fulfil all of the following:

- 1 = Unclear or Yes

- 2 = Unclear or Yes
- 4 = Unclear or Yes

- At least one of 5a, 5b, 5c = Yes

Reports that do not fulfil criteria for HIGH or MEDIUM are classified as LOW.

Appendix S4. Statistical methods for pooled estimates of reported causes

In general, the goal of a meta-analysis is not only to report the pooled estimate, but also to report how the results in the various individual studies are dispersed about the pooled estimate. One standard measure of dispersion (heterogeneity) in a meta-analysis is I^2. As pointed out by Higgins, I^2 is not an absolute measure of dispersion, but the proportion of total (observed) variation in the point estimates that is attributable to between-study variation[5]. For meta-analyses of cohort/observational studies, I^2 might not be particularly informative because the sample size is large and therefore the within-study variation is small; that is, almost all of the observed variation is between-study variation. For this present meta-analysis, the number of stillbirths in the identified studies was large, with a long tail to the right (median=300 stillbirths, mean=14670, inter-quartile range: 140, 1496). Unsurprisingly, I^2 for each of the individual causes-of-stillbirth was >90% (available on request); and, therefore not particularly informative. Another measure of dispersion/variation is $\tau 2$. However, a direct public-health interpretation of $\tau 2$ can be difficult; especially, as if often the case for meta-analyses, the analysis is on a transformed scale[6]. Therefore, we report 95% prediction intervals[6-8]. Prediction intervals are different from (e.g., typically wider than) confidence intervals and provide a direct measure of dispersion on the same scale the point estimates. The prediction interval tells us that if we were to select a hypothetical study at random from the same hypothetical universe of studies as those in the meta-analysis; then, in 95 of 100 hypothetical studies, the true outcome of interest in that study would fall in the range given by the prediction interval. Wide prediction intervals therefore tell us that different studies have reported widely different point estimates.

References

1. World Bank Group. World bank country and lending groups, country classification: World Bank Group; 2017 [cited 2017 18th May]. Available from: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups.
2. Blencowe H, Cousens S, Jassir FB, Say L, Chou D, Mathers C, et al. National, regional, and worldwide estimates of stillbirth rates in 2015, with trends from 2000: A systematic analysis. Lancet Glob Health. 2016;4(2):e98-e108.
3. Leisher SH, Teoh Z, Reinebrant H, Allanson E, Blencowe H, Erwich JJ, et al. Classification systems for causes of stillbirth and neonatal death, 2009–2014: An assessment of alignment with characteristics for an effective global system. BMC Pregnancy Childbirth. 2016;16(1):269.
4. Joanna Briggs Institute. The Joanna Briggs Institute critical appraisal tools for use in JBI systematic reviews, checklist for prevalence studies. Joanna Briggs Institute, 2016.
5. Higgins JP. Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. Int J Epidemiol. 2008;37(5):1158-60.
6. IntHout J, Ioannidis JPA, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. BMJ Open. 2016;6(7).
7. Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. J R Stat Soc Ser A Stat Soc. 2009;172(1):137-59.
8. Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. BMJ. 2011;342.