

Classification of maltreatment-related mortality by Child Death Review teams: How reliable are they?

Jared W. Parrish^{a,b,*}, Patricia G. Schnitzer^c, Paul Lanier^d,
Meghan E. Shanahan^{b,e}, Julie L. Daniels^e, Stephen W. Marshall^{b,e}

^a Alaska Division of Public Health, Section of Women's, Children's, and Family Health 3601CSt., Suite 358, Anchorage, AK, United States

^b The University of North Carolina at Chapel Hill, Injury Prevention Research Center, 137 East Franklin St, CB# 7505, Chapel Hill, NC 27599, United States

^c The National Center for the Review & Prevention of Child Deaths, c/o Michigan Public Health Institute, 1115 Massachusetts Ave. NW, Washington DC 20005, United States

^d The University of North Carolina at Chapel Hill, School of Social Work, 325 Pittsboro St. #3550, Chapel Hill, NC 27516, United States

^e The University of North Carolina at Chapel Hill, School of Public Health, 2101 McGavran-Greenberg Hall, CB# 7435, Chapel Hill, NC 27599, United States

ARTICLE INFO

Article history:

Received 21 June 2016

Received in revised form

29 December 2016

Accepted 7 March 2017

Available online 30 March 2017

Keywords:

Child Death Review

Maltreatment classifications

Maltreatment definitions

Inter-rater reliability

Child maltreatment

ABSTRACT

Accurate estimation of the incidence of maltreatment-related child mortality depends on reliable child fatality review. We examined the inter-rater reliability of maltreatment designation for two Alaskan Child Death Review (CDR) panels. Two different multidisciplinary CDR panels each reviewed a series of 101 infant and child deaths (ages 0–4 years) in Alaska. Both panels independently reviewed identical medical, autopsy, law enforcement, child welfare, and administrative records for each death utilizing the same maltreatment criteria. Percent agreement for maltreatment was 64.7% with a weighted Kappa of 0.61 (95% CI 0.51, 0.70). Across maltreatment subtypes, agreement was highest for abuse (69.3%) and lowest for negligence (60.4%). Discordance was higher if the mother was unmarried or a smoker, if residence was rural, or if there was a family history of child protective services report(s). Incidence estimates did not depend on which panel's data were used. There is substantial room for improvement in the reliability of CDR panel assessment of maltreatment related mortality. Standardized decision guidance for CDR panels may improve the reliability of their data.

1. Introduction

Child maltreatment, which includes both abuse and neglect, is a major public health problem (Kendall-Tackett, 2002). Reliable identification of fatalities from child maltreatment is critical for ongoing monitoring of this issue. Unfortunately, vital statistics data for children less than 16 years of age underrepresent maltreatment deaths by up to 50% (Crume, DiGuseppi, Byers, Sirotiak, & Garrett, 2002; Ewigman, Kivlahan, & Land, 1993; Herman-Giddens et al., 1999). Compared to vital statistics alone, the National Child Abuse and Neglect Data System (NCANDS) provides a more accurate assessment of the number of child maltreatment deaths each year. In fiscal year 2014, NCANDS estimated 1580 maltreatment related fatalities (US

* Corresponding author at: 3601CSt., Suite 358, Anchorage, AK 99503, United States.
E-mail addresses: jared.parrish@alaska.gov, jwp22@live.unc.edu (J.W. Parrish).

Department of Health and Human Services, Administration for Children and Families, Administration on Children, Youth and Families, Children's Bureau, 2016). However, even NCANDS undercounts child maltreatment deaths, due to variation in state classifications, non-standard definitions, voluntary state reporting, and reliance on child protection data (US, 2011).

To improve fatal maltreatment case detection and surveillance a range of methods have been evaluated. These include multi-source data (Putnam-Hornstein, Wood, Fluke, Yoshioka-Maxwell, & Berger, 2013; Schnitzer, Slusher, & Van Tuinen, 2004), hospital reports (Hampton & Newberger, 1985), capture re-capture (Palusci, Wirtz, & Covington, 2010), population survey (Finkelhor, Turner, Shattuck, & Hamby, 2013; McCurdy & Daro, 1994; Sedlak et al., 2010), and consensus or expert panel review (Leventhal, 1999; Palusci & Covington, 2014; Webster, Schnitzer, Jenny, Ewigman, & Alario, 2003). Some research however, suggests employing a public health model within the context of the CDR is the optimal approach for maltreatment detection and classification (Palusci et al., 2010; Schnitzer, Gulino, & Yuan, 2013).

Developed in 1979, the interagency CDR model was specifically designed to improve the identification of child deaths due to maltreatment (Durfee & Gellert, 1992). Unlike vital statistics death records or NCANDS, the CDR model (Covington, 2011) uses multiple sources of information and a multidisciplinary consensus review process to adjudicate the available information. Currently, the core interagency members of CDR include law enforcement, child protection, prosecutor/district attorney, medical examiner/coroner, public health, medical providers, and emergency medical services (Covington, Foster, & Rich, 2005). CDR panels typically classify potential maltreatment deaths into four categories: Yes, Probable, No, and Unknown. The use of these or other similar categories are widely used in fatal and nonfatal child maltreatment surveillance and research (Schnitzer, Slusher, Kruse, & Tarleton, 2011; Shanahan, Zolotor, Parrish, Barr, & Runyan, 2013).

Researchers have recommended that CDR teams utilize a systematic approach with standardized criteria when making maltreatment classifications to ensure consistency in classification (Palusci et al., 2010; Schnitzer et al., 2013). However, some evidence suggests considerable disagreement in the application of neglect classifications among and between CDR team members (Schnitzer, Covington, & Kruse, 2011; Schnitzer, Covington, & Kruse, 2011). Although the CDR model was originally developed to increase identification of child deaths due to maltreatment (Durfee & Gellert, 1992) this process has undergone minimal scientific scrutiny. It is currently unknown whether this process, used in nearly all 50 states, produces reliable maltreatment classifications. Research in other areas of death review (such as designation of preventability) has documented high levels of variability in the absence of clear methods to guide classification (MacKenzie, Steinwachs, Bone, Floccare, & Ramzy, 1992; McDermott, Cordner, & Tremayne, 1997).

The purpose of this study was to quantify the reliability of maltreatment classifications made through CDR consensus review in one state. A secondary purpose was to examine the effect of between-panel variation on incidence estimates for child maltreatment mortality. We selected Alaska for this research because their CDR team was readily accessible to us and interested in partnering.

2. Methods

We assessed inter-rater reliability of abuse, neglect, and negligence classification between two CDR panels (hereafter referred to as Panel 1 and Panel 2).

2.1. Alaska Maternal Infant Mortality Review – Child Death Review (MIMR-CDR)

Since 2008, the Alaska Maternal Infant Mortality Review – Child Death Review (MIMR-CDR) has used a broad definition to guide panels in classifying maltreatment-related mortality. The MIMR-CDR program defines Abuse as overt actions that cause harm, potential for harm, or threat of harm, Neglect as failure to provide for a child's physical or emotional needs or to protect from harm or threat of harm, and Negligence as failure to exercise reasonable care that would be expected of any other person in a similar situation. The operation of these definitions were modeled after those created by the National Center for the Review and Prevention of Child Deaths (NCRPCD) (The National Center for the Review and Prevention of Child Deaths, 2015) and other sentinel research (Schnitzer, Covington, Wirtz, Verhoek-Oftedahl, & Palusci, 2008). The MIMR-CDR is a public health program targeted with identifying patterns for prevention through systematic data collection, review, and recommendation. The determinations of these teams are for public health purposes only and are not used for individual criminal or child welfare proceedings.

Prior to review, the CDR administrative team collects, centralizes, and compiles comprehensive information from a variety of sources. Although the amount and breadth of information varies with each death, the core set of information often include: Medical Examiner autopsy and drug toxicology reports, death investigator reports, medical records (both child and mother if applicable), child protective services, law enforcement, village public safety records, first responder, Medicaid, publically available criminal justice records, and other relevant information if applicable.

The MIMR-CDR review process consists of three phases: 1) primary and secondary review, 2) case presentation and discussion, and 3) consensus classification. During the primary and secondary review, two committee members read through the case history file and take notes to document relevant information and circumstances of the death. After all deaths receive both a primary and secondary review, each is presented to the full panel followed by a discussion on the most probable causes, contributors, and preventability of each death. The discussion culminates in the committee making consensus classifications.

The MIMR-CDR program adapted the acts of omission and commission definitions and NCRPCD data elements to meet the needs of the Alaska program (The National Center for the Review and Prevention of Child Deaths, 2015). During this data

Table 1
Operational Definitions of Maltreatment Categories.

Category	Definition
Yes	Clear evidence and documentation through medical, child protection, law enforcement or medical examiner records that indicate maltreatment caused or contributed to the death
Yes Probably	Not enough evidence to make a definitive assessment with potentially conflicting information. Clear demonstration of unusual, questionable and/or suspicious factors present that strongly suggest maltreatment caused or contributed to the death
Unknown suspicious	No documentation or lack of information from sources indicating maltreatment caused or contributed to the death but multiple known risk factors or other suspicious evidence is deduced from the record leading to suspicion.
Unknown unlikely	No documentation or lack of information from sources indicating maltreatment caused or contributed to the death and no known risk factors or other suspicious evidence is deduced from the record.
No	Clear evidence and documentation through medical, child protection, law enforcement or medical examiner records that maltreatment did not cause or contribute to the death. No suspicious or known risk factors present or clearly explained how they are not associated with the death

collection period the Alaska maltreatment data elements for each Abuse, Neglect, and Negligence are classified as “Yes”, “Yes Probably”, “Unknown committee suspicious”, “Unknown but Unlikely”, “No”, whereas the NCRPCD data collection tool only allowed for this type of classification for the initial screening question of any act of omission or commission. Questions 1 and 2 of Section I in the NCRPCD data collection tool and the Alaska questions are available online as supplemental files (Appendix 1). [Table 1](#) outlines the criteria used by MIMR-CDR for operationalizing the maltreatment classifications ([Leeb, Paulozzi, Melanson, Simon, & Arias, 2008](#)).

2.2. Data collection

All infants and children (ages 0–4 years) born between 1/1/2009 and 12/31/2010 who died in Alaska prior to 12/31/2014 were eligible for inclusion in this study (n = 116). The initial review of these deaths by Panel 1 occurred during 2009–2014. Panel 2 met during 2015, with these 116 deaths randomly included with other reviews during the year. The maltreatment definitions and process used for data collection and review were consistent for both panels one and two. Panel members often participate in multiple review panels each year, with an average panel size being five reviewers (ranging from three to twelve). These reviews are conducted under a state medical review statute requiring membership to be 75% medical professionals. Partners are often invited to each meeting to expand the expertise at each review and often include child welfare, tribal, law enforcement, medical examiner, and mental health representatives. In the current study, we blinded the reviews of the second panel by removing all previous notes from the first panel. For some deaths that were particularly tragic or may have involved multiple professionals the second panel reviewers recalled the case. It is common practice however, for cases to be brought back to CDR teams for additional review, in the instances of this recall, the CDR administrator was instructed to treat these cases as an additional review and not disclose that they were part of a special study. Where possible we also ensured that, for each death, the primary and secondary reviewers for the second panel were unique from the first panel (this was not possible for two deaths). Six of the 39 total reviewers participated on both the initial and second panels accounting for 24 of the 101 deaths. For 37 deaths only the primary and secondary reviewers were recorded. Deaths not reviewed by the first panel prior to January 31, 2014 (n = 15) were excluded.

2.3. Statistical analyses

We assessed the inter-rater reliability between two review panels for the same set of deaths using weighted Kappa statistics, case-3 Intra-class correlation (ICC), and percent agreement. We calculated the weighted Kappa proposed by Fleiss and Cohen that uses a quadratic weighting of the discordant pairs to emphasize close matches more heavily than distal matches ([Fleiss, Cohen, & Everitt, 1969](#)). Overall agreement was derived by combining the three cross-classification tables of abuse, neglect, and negligence. Cell totals were created by averaging table cells to preserve the combined marginal distributions and prevent variance inflation.

We also assessed each classification matrix (Abuse, Neglect, Negligence) for rater bias using a rater bias coefficient and tested for proportional differences using Chi Square test ([Bishop, Fienberg, & Holland, 2007](#)). We calculated the incidence estimates and 95% Confidence Intervals (95% CI) for both panels, intersect between panels, and union of the two panels.

We examined the effect of the observed reliability on cumulative incidence estimates per 1000 live births. As a working assumption, the “Yes” and “Yes probably” categories were combined and assumed to represent a maltreatment positive surrogate. A lower bound on incidence was assumed to be the deaths classified as “Yes” or “Yes probably” by both panels (“panel-intersection”), and an upper bound on incidence was assumed to be the deaths classified as “Yes” or “Yes probably” by either panel (“panel-union”).

An independent abstractor conducted chart reviews and classified missing information as present or absent to test if discordance between panels was associated with missing information. We assessed the influence of demographic, social, and other known factors on discordance through prevalence ratios. We created two variables using two questions from the completed data extraction form. The first summarized each panel’s impression on “adequacy” of the information to

Table 2
Agreement classification between two review panels.

	Distribution of percent agreement by level (%)					Agreement ^a (%)	Kappa ^b (95% CI)	ICC ^c (95% CI)
	No	Unknown unlikely	Unknown suspicious	Yes Probably	Yes			
Abuse	65.0	11.1	35.3	37.5	77.8	69.3	0.70 (0.55, 0.85)	0.70 (0.59, 0.79)
Neglect	68.1	25.0	16.7	33.3	26.7	64.4	0.59 (0.42, 0.76)	0.59 (0.45, 0.71)
Negligence	63.1	18.8	15.8	27.3	36.7	60.4	0.53 (0.37, 0.70)	0.54 (0.38, 0.66)
Overall	65.3	15.8	22.2	30.0	36.8	63.4	0.60 (0.44, 0.77)	0.61 (0.53, 0.67)

Ref: ICC = Inter-Class Correlation; 95% CI = 95% Confidence Interval.

^a Total cross-diagonal agreement for all multi-categorical levels for combined panel summary tables.

^b Weighted Kappa statistic using the Fleiss-Cohen weights for overall cross-diagonal agreement.

^c ICC assuming a fixed set of k panels for a single target to assess oval cross-diagonal agreement.

determine cause of death, and the second captured the panel's impression whether caregiver substance use contributed to the death under review. For both variables, dichotomous classifications were created to code whether or not the panels agreed on these factors.

2.4. Statistical power

During planning for this study, the MIMR-CDR program imposed a projected limit of 100 deaths available for review. We calculated the study power to detect two-sided discordant differences ($\alpha = 0.05$) between observed and hypothetical kappa ranges of 0.4 and 0.9 with a fixed sample of 100 cases (Flack, Affi, Lachenbruch, & Schouten, 1988). At 80% power we could detect difference >0.1 for all null values above 0.4, and >0.2 at 0.4.

2.5. Institutional Review Board

The study was approved by the University of North Carolina at Chapel Hill non-biomedical Institutional Review Board. The MIMR-CDR program is conducted under the authority of the Alaska Division of Public Health as a public health initiative and on-going quality improvement process.

3. Results

Among the children born in Alaska between 1/1/2009 and 12/31/2010, 116 deaths subsequently occurred between 1/1/2009 and 12/31/2014 in Alaska. Of these, 87.1% ($n = 101$) were reviewed by the two MIMR-CDR panels. All 15 excluded deaths were classified as "natural" on the death certificate and 14 of the 15 occurred within the first month of life, with the infant having never left the hospital.

Among the reviewed deaths, the mean age at death was 8.9 months (range 0–58 months), 54% were male, 39% resided in Anchorage, and 67% were eligible for Medicaid. Among mothers of reviewed deaths, 39% were American Indian/Alaska Native, 50% were unmarried at delivery, and the mean maternal age at delivery was 25.6 years (range 15–42 years). Sex of child, residence, and maternal age were similar to the population of all Alaska births during 2009 and 2010. Deaths among Medicaid eligible, American Indian/Alaska Native race, and unmarried mothers were respectively 1.2, 1.3, and 1.4 times as high as all Alaska births.

3.1. Rater agreement

Agreement was generally higher for determinations of physical abuse than for neglect or negligence (Table 2). Percent agreement ranged from a low of 60.4% for negligence to a high of 69.3% for abuse (Table 2). Kappa and ICC scores followed a similar pattern. Summary percent agreement across all maltreatment categories was 63.4% with a weighted Kappa of 0.60 (Table 2).

Percent agreement was very low for most of the specific classification levels (Table 2). Only the "No" category for abuse, neglect, and negligence, and the "Yes" category for abuse, had percent agreement above 60%. Percent agreement for all other response categories, for all three outcomes, was below 40%. Eleven instances of extreme disagreement (yes vs no; representing complete opposite end of the scale) were documented (negligence = 8, neglect = 3, and abuse = 0).

3.2. Data element variation

We re-grouped the categories "Yes", "Yes Probably", "Unknown committee suspicious", "Unknown but unlikely", and "No" and created multiple sets of cross-classification summary tables based on different combinations to test for improved reliability. However no combination significantly improved agreement (data not shown).

3.3. Demographic comparisons

The pattern of agreement between panels for child and maternal characteristics was similar by sex, race, maternal smoking, and Medicaid eligibility at birth. The weighted Kappa was lower among infant (<365 days of age) compared to child (1 or more years of age) deaths (0.48 vs 0.77; $p=0.001$), unmarried mothers compared to married mothers at birth (0.47 vs 0.74; $p=0.005$), and for infants with no father's name listed on the birth certificate compared with infants with father's name listed (0.42 vs 0.67; $p=0.038$). Deaths listed as "natural" had the lowest overall agreement, followed by "accidental", "could not be determined", then "homicide" (Kappa = 0.47, 0.63, 0.71, 0.88, respectively).

3.4. Discordant classification description

Each death is evaluated for abuse, neglect, and negligence by the CDR; therefor multiple maltreatment classifications can be applied to the same case. To better understand the potential influence of case information on overall classification we assessed the 49 deaths with any disagreement between panel 1 and panel 2. Almost half (44.9%) were discordant for each of the three maltreatment classifications, 28.6% for two, and 26.5% for only one.

Among the 77 infant deaths, 30 neonates (0–28 days) were reviewed and 11 (36.6%) were discordant for any maltreatment, 47 post-neonatal (29364 days) were reviewed and 28 (59.6%) were discordant for any maltreatment. Among the 34 child deaths (356+ days), 10 (41.6%) were discordant for any maltreatment.

Out-of-hospital deaths that occurred in the sleep environment accounted for 68.1% ($n=32$) of all post-neonatal deaths, with 71.9% ($n=23$) having at least one discordant maltreatment classification. Among neonatal deaths, sleep environment deaths accounted for 26.7% ($n=8$), with 75.0% ($n=6$) discordant in maltreatment designation between panels.

Discordance was more likely among children ages 12–23 months compared to ages 24+ months (53.8% vs. 27.3%). The mechanism of death was clear for 60% of the deaths with at least one discordant classification. Caregiver responsibility was attributed to 66.7% ($n=16$) of all child deaths, and 80% ($n=8$) among those discordant.

Among child characteristics investigated (sex, age, race, Medicaid eligibility, and residence), only rural residence was strongly associated with discordance classification between panels for abuse ($p=0.009$), negligence ($p=0.037$), and neglect ($p=0.068$) when compared to urban residence (urban included Anchorage, Mat-Su, Juneau, and Fairbanks census boroughs, and rural all other census boroughs).

Discordance between panels was associated with unmarried maternal marital status for abuse, neglect and negligence, maternal smoking during pregnancy for abuse and negligence, out-of-hospital deaths, and family history of child protection for abuse and neglect only (Table 3). Discordance increased when the panels disagreed on the adequacy of information available for assigning the cause of death, and further if they disagreed on if caregiver substance use was related to the death (Table 3). Panels tended to agree that adequate information was available for urban deaths more often compared to deaths of rural residents (75.8% vs 61.5%; $p=0.321$).

Table 3

Proportion ratio (PR) for the association between demographics/review information and discordant classifications between review panels ($n=101$).

	N (101)	Abuse		Neglect		Negligence	
		PR (95% CI)	p-value	PR (95% CI)	p-value	PR (95% CI)	p-value
Child Characteristics							
Sex (male vs female)	101	1.2 (0.7, 2.2)	0.548	1.0 (0.6, 1.6)	0.919	0.8 (0.5, 1.3)	0.340
Infant Death (<24mo vs 24+ mo)	101	1.6 (0.7, 3.8)	0.242	1.1 (0.6, 2.1)	0.801	1.5 (0.7, 2.9)	0.243
Race (Native vs not Native)	101	1.5 (0.8, 2.7)	0.191	1.6 (0.9, 2.7)	0.088	1.4 (0.9, 2.3)	0.146
Medicaid (eligible vs not)	94	1.5 (0.7, 3.2)	0.328	1.5 (0.7, 3.3)	0.261	1.2 (0.7, 2.2)	0.575
Residence (Urban/Rural)	101	2.2 (1.2, 3.9)	0.009	1.6 (1.0, 2.7)	0.068	1.7 (1.1, 2.7)	0.037
Family/Review Characteristics							
Marital Status (Unmarried vs Married)	95	3.2 (1.4, 7.2)	0.001	3.3 (1.6, 6.8)	0.000	2.6 (1.4, 5.0)	0.001
Maternal smoking (any vs none)	94	1.7 (0.9, 3.2)	0.090	1.5 (0.9, 2.5)	0.180	1.6 (1.0, 2.6)	0.079
Family history of CPS reports (yes vs no)	95	2.4 (1.4, 4.1)	0.004	2.1 (1.3, 3.5)	0.006	1.4 (0.9, 2.3)	0.190
Out of hospital Death (Yes vs no)	100	2.6 (1.2, 5.8)	0.008	2.6 (1.3, 5.4)	0.003	3.0 (1.5, 6.0)	<0.001
Information adequate (Panels agreed vs did not agree) ^a	94	2.4 (1.4, 4.1)	0.003	2.1 (1.3, 3.3)	0.006	1.8 (1.2, 2.9)	0.019
Caregiver substance use contributed (Panels agreed vs did not agree) ^b	101	2.1 (1.2, 3.6)	0.018	2.1 (1.3, 3.4)	0.009	1.7 (1.1, 2.7)	0.034

Level of discordance for abuse was 31/101 (30.7%); Neglect 36/101 (35.6%); Negligence 40/101 (39.6%)

^a This was determined by comparing both panels response to the question "Was the information available for review adequate for the committee to determine the cause(s) of death?" We created a dichotomous variable where 0 = panels agreed the information was adequate or not, and 1 = panels did not agree on adequacy of information. "Information" refers to all the records about the death available at the time of the review.

^b This was determined by comparing both panels response to the question "Did substance use by someone else contribute to the child's death?" We created a dichotomous variable where 0 = panels agreed substance use did or did not contribute, and 1 = panels did not agree.

3.5. Review panel

Changes in MIMR-CDR protocol in recording panel member attendance impacted our ability to adequately evaluate panel member composition. For 64 of the 101 deaths we had complete reviewer information for both panels. Among these 64 deaths, no significant difference in discordance was detected for committee composition, number of members present, or sex of the panel members; however our power to detect differences was minimal (data not shown).

3.6. Impact of classification variation on maltreatment-related mortality estimates

Incidence per 1000 live births of neglect and negligence documented large variations between the panel-intersection and panel-union estimates (0.44 vs 1.10 and 0.79 and 1.63, respectively); however, abuse estimates were much more stable (0.44 vs 0.75). Both panels independently identified statistically equivalent incidence estimates for all three maltreatment classifications, with abuse having the lowest incidence, followed by neglect, then negligence (Fig. 1).

4. Discussion

We detected that for any fatal maltreatment classification (abuse, neglect, or negligence), the inter-rater reliability between MIMR-CDR panels was in the “moderate” to “substantial” range of agreement (Landis & Koch, 1977). Given the potential public health implications resulting from CDR findings, our opinion is that this level of precision is less than optimal to adequately inform and evaluate public health policy and prevention interventions to reduce child maltreatment-related mortality.

Both panels independently produced nearly equivalent 5-year incidence estimates for fatal maltreatment. We caution that consistency of estimation is not the same as validity of estimation. The “true” population incidence of abuse, neglect, and negligence remains unknown. Based on results from only two panels, and assuming the “true” incidence likely ranges somewhere between panel intersect and panel union estimates (Fig. 1), the 5-year incidence proportion could be as low as 1.2 and as high as 2.0 per 1000 live births for any maltreatment. For abuse the range was 0.44–0.75, for neglect the range was 0.44–1.1, and for negligence the range was 0.8–1.6. The ranges in these estimates reflect variance in the identification and interpretation of known demographic and social factors, adequacy of information, variation in panel membership, and likely misinterpretation or poor operation of definitions used for classification. Until these “influences” are consistently interpreted or mitigated by CDR panels, overall and subgroup incidence estimates will likely continue to be highly variable.

Our findings are consistent with a previous study assessing presence/absence of demographic, social, and service based attributes that influence individual CDR panel member death classification. Using 20 vignettes of unintentional injury death,

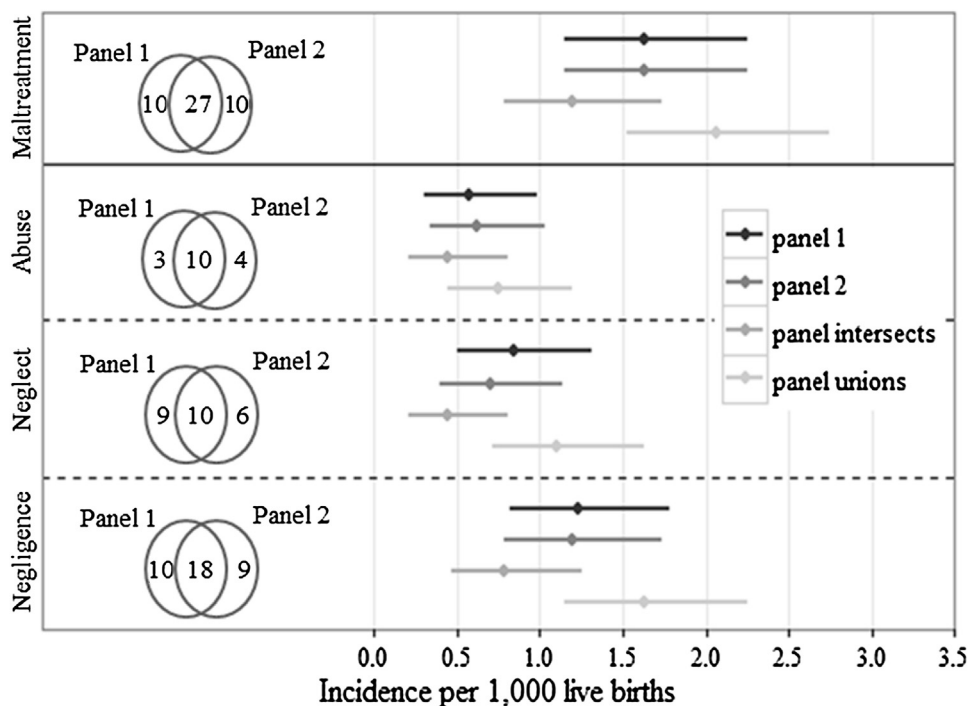


Fig. 1. Maltreatment-related (Yes/Yes probably) mortality classification incidence estimation comparison by panel reviews, union and intersect, 2009–10 Alaska birth cohort (n = 22,707 live births).

this prior study documented high levels of variability within and between CDR members in assigning caregiver responsibility (Schnitzer, Slusher et al., 2011; Schnitzer, Covington et al., 2011). Similarly, we found that discordant classifications of maltreatment between panels tended to be more common when certain “leading factors” such as unmarried mother, no father on the birth certificate, maternal smoking, or history of family CPS involvement were detected. In the vignette-based study, when these factors were introduced into the scenarios, they strongly influenced neglect classification for many of the panel members. The magnitude of the influence of these identified factors varies between panel members (Schnitzer, Slusher et al., 2011; Schnitzer, Covington et al., 2011), consistent with the recommendation that clear guidance is needed to assist CDR panel members in the interpretation of these leading factors when assessing the potential for child maltreatment (Schnitzer et al., 2013).

Completeness of information available for review can also influence committee determinations. We found large discordance in maltreatment classification when panels disagreed on the adequacy of information for making a classification and if caregiver substance use was associated with the death. The synthesis and extrapolation of information by CDR teams to influence classification was especially evident in the eleven instances of extreme disagreement. The majority of these extreme cases were disagreement in negligence and the influence of more distal information or detection of missed information (such as history of substance abuse and maltreatment). These data suggest that panels processed information differently, particularly for out-of-hospital infant deaths occurring in the sleep environment. Efforts should be directed at developing guidelines in assigning caregiver responsibility, especially among sleep-related deaths that consider the influence of factors such as environmental hazards, substance use, family social welfare, and law enforcement contacts. A large proportion of disagreement between panels could be mitigated with increased attention in this single area, which could be adapted into the current CDC guidance for classifying Sudden Unexplained Infant Death (SUID) (Shapiro-Mendoza et al., 2014).

These findings address a gap in existing knowledge about classifying maltreatment fatalities within the context of the CDR process. It appears that the operational definitions, at least as implemented by the Alaska CDR panels, do not provide clear direction on how to systematically interpret missing information or identified prominent factors. The National Center for the Fatality Review and Prevention is establishing national standards and a common data collection process, but the impact of this work will be limited if CDR teams fail to implement these standards consistently (Fraser, Sidebotham, Frederick, Covington, & Mitchell, 2014).

Reliability of public health surveillance data directly impacts the ability to take action to improve health (German et al., 2001). The lack of confidence in the true incidence of maltreatment fatalities creates challenges for those tracking state-level trends to inform and assess programs and policies. Precise estimates are needed for efficient program and resource planning and also to accurately detect the impact of population-level interventions and policies that seek to reduce the incidence of CAN-related deaths. Given that the purpose of the Child Death Review process is to inform action to prevent future fatalities, it is important that the deaths are accurately and consistently classified over time. Doing so will ensure that the recommendations that arise from the Child Death Review are appropriately focused in order to achieve a measurable public health impact.

4.1. Limitations

The findings of the study are subject to limitations. First, this study was limited to approximately 100 cases and this impacts the precision of these data. Second, the MIMR-CDR membership is predominantly medical personnel with limited contributions from other agencies. The multidisciplinary CDR process is likely most effective with well-balanced representation from multiple sectors. Third, a number of MIMR-CDR reviewers participated in both the initial and second panel review which could impact the reliability of classifications for some cases. We attempted to mitigate this influence by ensuring the primary and secondary reviewers were unique from the initial review, removed all notes and documentation, and attempted to have at least 3 months laps prior to a second review (for over 90% of the deaths at least 1 year elapsed). Fourth, the completeness of information has improved over time, thus more recent cases often have more comprehensive case files. Finally, the Alaska review process used a state-specific tool that may not be generalizable to the national case reporting system.

5. Conclusion

For over a decade, much research described and documented the underascertainment of maltreatment fatalities in official sources (Herman-Giddens et al., 1999), and suggested novel approaches for expanding these estimates (such as data integration and capture re-capture) (Palusci et al., 2010; Schnitzer et al., 2013). However, variation in definitions, processes, and information continue to contribute to measurement uncertainties (Putnam-Hornstein et al., 2013). Research suggests the best source for measuring maltreatment mortality is the CDR. Our study provides the first known reliability assessment of CDR panel classification of maltreatment mortality and provides evidence for the need to develop and validate a set of guidelines or decision process for classifying maltreatment, especially neglect and negligence.

Valid and reliable surveillance data is the cornerstone of the public health prevention model. CDR is a key process for improving maltreatment fatality estimates. Additional research is needed to assist in developing systematic processes for making consistent classifications. Future research should also assess the impact of committee composition on classifica-

tion designations. Until we can increase the consistency in classifying maltreatment deaths, incidence estimates will vary unpredictably and reduce our ability to implement and measure the impact of prevention efforts over time.

Conflict of interest statement

At the time of the study the corresponding author was both a doctoral student at the University of North Carolina and an employee of the Alaska Division of Public Health (ADPH). The ADPH requires all manuscripts to be reviewed by an executive leadership committee. No other potential conflicts of interest are noted.

Financial disclosure

The research was supported in part by the “Travis Fund”. This fund, established by a private donor in Alaska, supports the scholarly research efforts of a graduate student at the University of North Carolina to benefit the children of Alaska. Its primary mission is to support efforts to quantify and elucidate the complex causal mechanisms leading to maltreatment. The primary author was supported by this research assistantship during the study. S.W.M. and the University of North Carolina Injury Prevention Research Center are partially supported by an award from the National Center for Injury Prevention and Control, Centers for Disease Control and Prevention (R49/CE002479).

Acknowledgements

This work was only possible due to the contributions of the Maternal Infant Mortality and Child Death Review (MIMR-CDR) program team within the Alaska Division of Public Health. We acknowledge each volunteer committee member/reviewer of the MIMR-CDR and their time in reviewing these cases. This research was funded in part by the Title V Maternal Child Health Block Grant, and the Travis Fund.

References

- Bishop, Y. M., Fienberg, S. E., & Holland, P. W. (2007). *Discrete multivariate analysis: Theory and practice*. Springer Science & Business Media.
- Covington, T., Foster, V., & Rich, S. (2005). *A program manual for child death review*. Okemos, MI: The National Center for Child Death Review.
- Covington, T. M. (2011). The US national child death review case reporting system. *Injury Prevention, Journal of the International Society for Child and Adolescent Injury Prevention*, 17(Suppl. (1)), i34–i37. <http://dx.doi.org/10.1136/ip.2010.031203>
- Crume, T. L., DiGuseppi, C., Byers, T., Sirotiak, A. P., & Garrett, C. J. (2002). Underascertainment of child maltreatment fatalities by death certificates, 1990–1998. *Pediatrics*, 110(2 Pt. 1), e18.
- Durfee, M. J., & Gellert, G. A. (1992). Origins and clinical relevance of child death review teams. *JAMA*, 267(23), 3172–3175.
- Ewigman, B., Kivlahan, C., & Land, G. (1993). The Missouri child fatality study: Underreporting of maltreatment fatalities among children younger than five years of age, 1983 through 1986. *Pediatrics*, 91(2), 330–337.
- Finkelhor, D., Turner, H. A., Shattuck, A., & Hamby, S. L. (2013). Violence, crime, and abuse exposure in a national sample of children and youth: An update. *JAMA Pediatrics*, 167(7), 614–621.
- Flack, V. F., Afifi, A., Lachenbruch, P., & Schouten, H. (1988). Sample size determinations for the two rater kappa statistic. *Psychometrika*, 53(3), 321–325.
- Fleiss, J. L., Cohen, J., & Everitt, B. (1969). Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin*, 72(5), 323.
- Fraser, J., Sidebotham, P., Frederick, J., Covington, T., & Mitchell, E. A. (2014). Learning from child death review in the USA, England, Australia, and New Zealand. *The Lancet*, 384(9946), 894–903.
- German, R. R., Lee, L., Horan, J., Milstein, R., Pertowski, C., & Waller, M. (2001). Updated guidelines for evaluating public health surveillance systems. *MMWR Recommendations and Reports*, 50(1–35)
- Hampton, R. L., & Newberger, E. H. (1985). Child abuse incidence and reporting by hospitals: Significance of severity, class, and race. *American Journal of Public Health*, 75(1), 56–60.
- Herman-Giddens, M. E., Brown, G., Verbiest, S., Carlson, P. J., Hooten, E. G., Howell, E., et al. (1999). Underascertainment of child abuse mortality in the United States. *JAMA*, 282(5), 463–467.
- Kendall-Tackett, K. (2002). The health effects of childhood abuse: Four pathways by which abuse can influence health. *Child Abuse & Neglect*, 26(6), 715–729.
- Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 363–374.
- Leeb, R. T., Paulozzi, L., Melanson, C., Simon, T., & Arias, I. (2008). *Child maltreatment surveillance: Uniform definitions for public health and recommended data elements. (Document No. 1.0)*. National Center for Injury Prevention and Control: The Center for Disease Control and Prevention.
- Leventhal, J. M. (1999). The challenges of recognizing child abuse: Seeing is believing. *JAMA*, 281(7), 657–659.
- MacKenzie, E. J., Steinwachs, D. M., Bone, L. R., Floccare, D. J., Ramzy, A. I., & Preventable Death Study Group. (1992). Inter-rater reliability of preventable death judgments. *Journal of Trauma and Acute Care Surgery*, 33(2), 292–303.
- McCurdy, K., & Daro, D. (1994). Child maltreatment A national survey of reports and fatalities. *Journal of Interpersonal Violence*, 9(1), 75–94.
- McDermott, F., Cordner, S. M., & Tremayne, A. (1997). Reproducibility of preventable death judgments and problem identification in 60 consecutive road trauma fatalities in Victoria, Australia. *Journal of Trauma and Acute Care Surgery*, 43(5), 831–839.
- Palusci, V. J., & Covington, T. M. (2014). Child maltreatment deaths in the US national child death review case reporting system. *Child Abuse & Neglect*, 38(1), 25–36.
- Palusci, V. J., Wirtz, S. J., & Covington, T. M. (2010). Using capture-recapture methods to better ascertain the incidence of fatal child maltreatment. *Child Abuse & Neglect*, 34(6), 396–402.
- Putnam-Hornstein, E., Wood, J. N., Fluke, J., Yoshioka-Maxwell, A., & Berger, R. P. (2013). Preventing severe and fatal child maltreatment: Making the case for the expanded use and integration of data. *Child Welfare*, 92(2), 59–75.
- Schnitzer, P. G., Slusher, P., & Van Tuinen, M. (2004). Child maltreatment in Missouri: Combining data for public health surveillance. *American Journal of Preventive Medicine*, 27(5), 379–384.
- Schnitzer, P. G., Covington, T. M., Wirtz, S. J., Verhoek-Oftedahl, W., & Palusci, V. J. (2008). Public health surveillance of fatal child maltreatment: Analysis of 3 state programs. *American Journal of Public Health*, 98(2), 296–303.

- Schnitzer, P. G., Gulino, S. P., & Yuan, Y. T. (2013). Advancing public health surveillance to estimate child maltreatment fatalities: Review and recommendations. *Child Welfare, 92*(2), 77.
- Schnitzer, P. G., Covington, T. M., & Kruse, R. L. (2011). Assessment of caregiver responsibility in unintentional child injury deaths: Challenges for injury prevention. *Injury Prevention: Journal of the International Society for Child and Adolescent Injury Prevention, 17*(Suppl 1), i45–i54.
- Schnitzer, P. G., Slusher, P. L., Kruse, R. L., & Tarleton, M. M. (2011). Identification of ICD codes suggestive of child maltreatment. *Child Abuse & Neglect, 35*(1), 3–17.
- Sedlak, A. J., Mettenburg, J., Basena, M., Peta, I., McPherson, K., & Greene, A. (2010). *Fourth national incidence study of child abuse and neglect (NIS-4)*. Washington, DC: US Department of Health and Human Services. Retrieved on July, 9, 2015.
- Shanahan, M. E., Zolotor, A. J., Parrish, J. W., Barr, R. G., & Runyan, D. K. (2013). National, regional, and state abusive head trauma: Application of the CDC algorithm. *Pediatrics, 132*(6), e1546–53.
- Shapiro-Mendoza, C. K., Camperlengo, L., Ludvigsen, R., Cottengim, C., Anderson, R. N., Andrew, T., et al. (2014). Classification system for the sudden unexpected infant death case registry and its application. *Pediatrics, 134*(1), e210–9.
- The National Center for the Review & Prevention of Child Deaths. (2015). *Child death review case reporting system: Data dictionary*. (No. V4.0). US Department of Health and Human Services, Administration for Children and Families, Administration on Children, Youth and Families, Children's Bureau. (2016). *Child maltreatment 2014*.. Available from <http://www.acf.hhs.gov/programs/cb/research-data-technology/statistics-research/child-maltreatment>
- U.S. government accountability office. (2011) *Child maltreatment: Strengthening national data on child fatalities could aid in prevention* (No. GOA-11-599).
- Webster, R. A., Schnitzer, P. G., Jenny, C., Ewigman, B. G., & Alario, A. J. (2003). Child death review: The state of the nation. *American Journal of Preventive Medicine, 25*(1), 58–64.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.chiabu.2017.03.003>.