Received: 12 July 2023 | Revised: 15 January 2024 | Accepted: 2 May 2024

DOI: 10.1111/birt.12833

ORIGINAL ARTICLE

BIRTH ISSUES IN PERINATAL CARE WILEY

An exploratory review on the empirical evaluation of the quality of reporting and analyzing labor duration

Emilienne Celetta MSc[1] | Loukia M. Spineli PhD[1] | Valérie Avignon MSc[2] | Hanna Gehling MSc[1] | Mechthild M. Gross PhD[1]

[1]Midwifery Research and Education Unit, Hannover Medical School, Hannover, Germany

[2]Lausanne University Hospital – CHUV, Lausanne, Switzerland

Correspondence
Mechthild M. Gross, Hebamme, Midwifery Research and Education Unit, Hannover Medical School (MHH), Carl-Neuberg-Str. 1, 30625 Hannover, Germany.
Email: gross.mechthild@mh-hannover.de

Abstract

Introduction: This exploratory review aimed to provide empirical evidence on the definitions of labor, the statistical approaches and measures reported in randomized controlled trials (RCTs) and observational studies measuring the duration of labor.

Methods: A systematic electronic literature search was conducted using different databases. An extraction form was designed and used to extract relevant data. English, French, and German studies published between 1999 and 2019 have been included. Only RCTs and observational studies analyzing labor duration (or a phase of labor duration) as a primary outcome have been included.

Results: Ninety-two RCTs and 126 observational studies were eligible. No definition of the onset of labor was provided in 21.7% ($n=20$) of the RCTs and 23.8% ($n=30$) of the observational studies. Mean was the most frequently applied measure of labor duration in the RCTs (89.1%, $n=82$), and median in the observational studies (54.8%, $n=69$). Most RCTs (83%, $n=76$) and observational studies (70.6%, $n=89$) analyzed labor duration using a bivariate method, with the *t*-test being the most frequently applied (45.7% and 27%, respectively). Only 10.8% ($n=10$) of the RCTs and 52.4% ($n=66$) of the observational studies conducted a multivariable regression: 3 (30%; out of 10) RCTs and 37 (56%; out of 66) observational studies used a time-to-event adapted model.

Conclusion: This survey reports a lack of agreement with respect to how the onset of labor and phases of labor duration are presented. Concerning the statistical approaches, few studies used survival analysis, which is the appropriate statistical framework to analyze time-to-event data.

KEYWORDS

labor duration, labor onset, midwifery, observational studies, randomized controlled studies, survival analysis

1 | INTRODUCTION

The appropriate assessment of labor duration may affect the care women and their babies receive. Determining normal labor duration makes it possible to analyze physiologic outliers in length. However, the threshold chosen to diagnose abnormal labor duration should be carefully investigated since it may justify using interventions to accelerate labor.[1–3] In order to address the current increase in labor interventions, a deeper understanding of the dynamics during labor progression is crucial.[4–10] Nevertheless, the significant methodological heterogeneity and reporting variations may limit the meta-analysis of papers.[11]

Numerous challenges have been reported when researching labor duration.[1,4,11] The appropriate definition of the onset of early or active labor, selecting and applying appropriate statistical approaches, reporting the relevant measures, and considering the impact of interventions and events determine the accuracy of results.[8,11–13] The current variation in approaches to labor duration is a barrier to comparing studies and suggesting comprehensible guidelines for performing new studies.[4,11,14,15]

The most frequently reported challenges when analyzing labor duration are the lack of standard definitions to diagnose the onset of labor and its different stages and phases.[4,9,11,12] The traditional division of labor has been used in the whole paper. Here, labor is distinguished into three stages, starting with the onset of labor and ending with the delivery of the placenta. The term "first stage of labour" refers to the period between labor onset and cervical dilatation of up to 10 cm.[13–15] This stage is subdivided into two phases: the latent phase or early labor and the active phase, both mainly characterized by cervical ripening and dilatation.[13] The second stage refers to the time between full dilatation of the cervix and birth.[13]

In addition to the complex and lack of standard definitions in labor, considerable heterogeneity in statistical approaches has been reported.[11,14] Prolonged birth processes result in a long right tail of the distribution, and thus, labor duration is inherently positively right-skewed.[6,16,17] Therefore, a median has been claimed as the appropriate measurement of the central tendency for labor duration[1,11,12,16]; however, many studies report only the mean, which may lead to an overestimation of the investigated labor duration.[11,14]

With respect to statistical analysis, different approaches have been pursued to analyze labor duration. Even if the outcome is inherently time-related, some authors do not use survival analysis.[12] Survival analysis encompasses a variety of statistical methods to handle time-to-event data correctly.[18] For instance, the Kaplan–Meier survival curve is a popular visualization method to represent the cumulative probability of giving birth in a certain length of time.[19,20] Several statistical tests and regression models of increasing complexity have been developed to analyze time-to-event data and to account for censored data that are unavoidable in clinical research.[21–23]

In addition, particular characteristics and interventions are associated with labor duration, including time-varying interventions, such as the artificial rupture of membranes or an epidural, but also women's characteristics, which are time-constant.[11,24] Time-varying predictors should be analyzed with appropriate statistical methods, such as regression models designed to deal with such time-to-event data.[20,25] For instance, Gross et al.[5] applied a Cox proportional hazards regression model, which adjusted for time-constant and time-varying confounders. Another example is Zhang et al.,[16] who analyzed their interval-censored data using an interval-censored regression.

The Cox proportional hazards regression has been the most widely used model for time-to-event data for having been built upon simple assumptions, namely linear association between the response variable in the logarithmic scale and the predictors and constant hazard ratio over time (the basis of this model). However, defending these assumptions can be challenging, giving rise to several more flexible methods as a viable alternative.[23] Interval-censored data constitute a special case of time-to-event data where the exact value is not observed; therefore, the interval around this value is considered, leading to information loss for not knowing the exact distribution of the variable within the intervals. Several approaches have been developed to handle interval-censored data, with interval-censored regression being such an approach.

Based on distribution assumptions of the survival time, different methods with different properties and interpretations can be used.[26,27] These methods are distinguished into the non-parametric methods that make no distribution assumption about the hazard function; semi-parametric methods, such as the Cox proportional hazards model, that consider a functional link between the response and predictors, but non-parametric methods to estimate the hazard function; and parametric models that make a specific assumption about the distribution of the hazards over time.[28] Examples of parametric survival models include exponential models that assume the hazards to remain constant over time; Weibull and Gompertz models that assume the hazards to increase or decrease monotonically; and log-logistic, log normal, and standardized gamma models that represent hazards as being monotonically decreasing or initially increasing and then decreasing on a turning point.[28] However, once there are multiple changes in the slope of the hazard function, the parametric models may not provide accurate results, and flexible models, such as fractional polynomials and splines, may be appropriate.[20]

Nevertheless, despite advances in statistical modeling of time-to-event data, linear or logistic regression models are still pursued when studying labor duration.[12] The former may lead to negative predicted time, where time is the dependent variable. The latter requires determining a threshold and categorizing the subjects to those with a time up to this threshold and those with a time exceeding the threshold and using this binary variable as the response in a binary logistic regression model.

The design quality, statistical analysis, and reporting of a study evaluating normal labor duration may implicate the decision-making for optimal midwifery care. Therefore, the present review aims to offer an exploratory empirical evaluation of the definition, measure, and statistical analysis of labor duration as reported in randomized controlled trials (RCTs) and observational studies.

2 | MATERIALS AND METHODS

2.1 | Eligibility criteria

The present study has been conducted following the guidelines for systematic reviews.[29] We included articles reporting labor duration in minutes or hours as the primary outcome regardless of the phase or stage of labor. Randomized controlled trials, non-RCTs, longitudinal and cross-sectional cohort studies published during the last 20 years were considered. Only studies in English, German, or French were reviewed. We excluded studies initially considering labor duration as a binary categorical variable, such as prolonged labor versus short labor. However, we included studies that measured labor duration as time-to-event data and then dichotomized it.

2.2 | Search strategy and selection process

Keywords such as "labor" or "labour" and "duration" or "length*" or "progress*" were inserted into Medline, Embase and Cochrane Central, CINHAL EBSCO, and Google Scholar. A strategy was developed combining free terms and terms from the thesaurus adapted to the specific databases. The systematic search was conducted between November and December 2019 in collaboration with experts from the Medical Library of Lausanne University Hospital and the University of Lausanne. The detailed search strategy is found in the Supporting Information (Appendix S1). Two reviewers (EC, VA) used the Covidence® software to double-screen in a standardized approach the titles and abstracts of the studies retrieved from the literature search. Relevant studies were screened in full-text, and reasons for exclusion were documented. A third reviewer (MMG) was consulted in case of conflict

2.3 | Data collection process

We devised extraction items documented to influence the quality of findings with respect to labor duration. These extraction items have been organized into four domains: (1) general study characteristics, such as publication year, sample size, study design, and sample characteristics; (2) definition of labor-related concepts; (3) descriptive statistics and statistical testing; and (4) multivariable analysis. The extraction form can be found in the Supporting Information (Appendix S2). One reviewer (EC) extracted all necessary information, and the second reviewer (VA) randomly selected and checked 10% of the extracted studies for potential errors. Disagreements were resolved by discussion between the reviewers. In case of disconcordance, the authors consulted a third reviewer. HG and MMG extracted the data from the German studies.

2.4 | Pilot study

Initially, we performed a pilot study to test the extraction form and ensure completeness in the necessary extraction items. The search was tested in two databases (Embase and CINAHL) with a restricted publication date of 2 years (2017–2019). The selection process and data extraction form were piloted on 20 randomly selected eligible studies.

2.5 | Data analysis

The extraction form consisted of 18 items. We used one table for each domain (four tables in total) to summarize the corresponding items for the whole sample and by study design (RCTs and observational studies). We used the median, interquartile range (IQR), and range to describe the continuous items and absolute and relative frequencies for the categorical items. The Statistical Package SPSS version 26 (IBM) software was used for the analysis.

3 | RESULTS

3.1 | Results of the search strategy and description of the sample

The search strategy retrieved 10,154 studies. Of those studies, 4645 (46%) were excluded for being duplicates, and 4931 (49%) were deemed ineligible according to the inclusion/exclusion criteria during title and abstract screening. Of the remaining 578 studies that underwent full-text assessment, 360 (62%) were excluded for not meeting our inclusion criteria. Overall, 218 studies were included in the present study (Figure 1). The list of included studies is found in Appendix S3.

3.2 | General study characteristics

Of 218 studies, we identified 92 RCTs (42.2%) and 126 observational studies (57.8%). The median sample size was 239 (range: 27–146,904) in the whole sample: RCTs had a comparatively smaller sample than the observational studies (median: 161 with range: 40–960, and median: 651 with range: 27–146,904, respectively). Most studies were conducted in hospitals ($n=90$, 97.8% RCTs and $n=122$, 96.8% observational studies). Two observational studies reported birth center or home as birth settings. Most RCTs focused on nulliparous women ($n=62$, 67.4%). By contrast, most observational studies ($n=81$, 64.3%) included both multiparous and nulliparous (Table 1).

3.3 | Definition of labor-related concepts

Both study designs investigated mostly the second stage (RCTs: $n=70$, 76.1% and observational studies: $n=88$, 69.8%) (Table 2). However, the first stage succeeded mostly in observational studies ($n=58$, 46.0%), followed by the whole labor process ($n=36$, 28.6%) and the active phase ($n=30$, 23.8%). Contrariwise, RCTs focused mostly on the active phase ($n=49$, 53.3%), followed by the first stage ($n=31$, 33.7%), and the whole labor process ($n=30$, 32.6%). The latent phase was underrepresented in both study designs (3.3% of RCTs and 6.3% of observational studies) (Table 2).

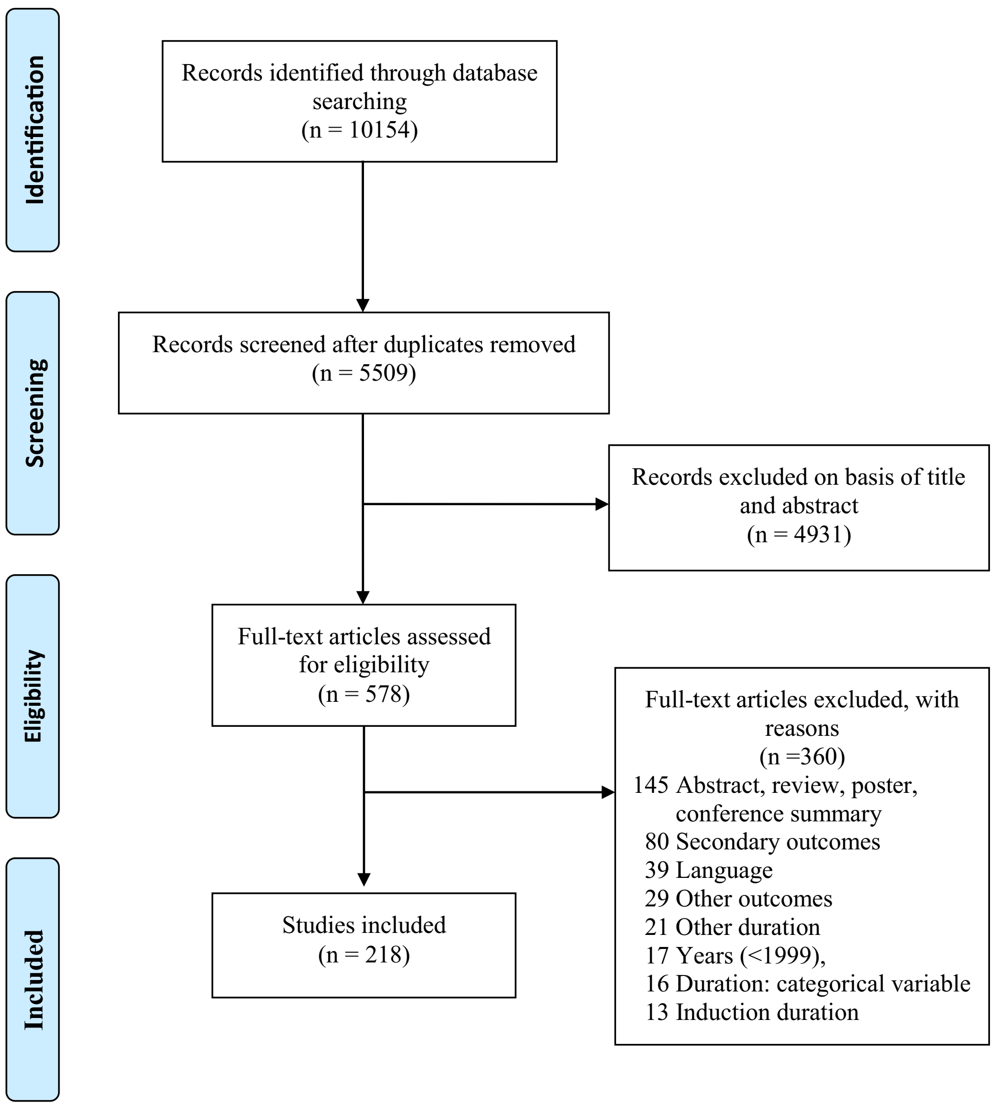

FIGURE 1 Flow chart of the process selection. [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Summary of study general characteristics.

Characteristics	Randomized controlled studies ($n=92$)	Observational studies ($n=126$)	Whole sample ($n=218$)
Year of publication			
Median (IQR; min–max)	2012 (2008–2016; 1999–2019)	2012 (2006–2016; 1999–2019)	2012 (2007–2016; 1999–2019)
Sample size			
Median (IQR; min–max)	161 (107–209; 40–960)	651 (200–3370; 27–146,904)	239 (132–1093; 27–146,904)
Period of the data record			
Median (IQR; min–max)	2010 (2007–2015; 1996–2017)	2008 (2004–2013; 1965–2018)	2008 (2005–2014; 1965–2018)
Missing cases, *n* (%)	19 (20.7)	16 (12.7)	35 (16.1)
Setting of birth *n* (%)			
Birth center/home birth	0 (0)	2 (1.6)	2 (0.9)
Hospital	90 (97.8)	122 (96.8)	212 (97.2)
Not specified	2 (2.2)	2 (1.6)	4 (1.8)
Parity of the sample *n* (%)			
Nulliparous	62 (67.4)	40 (31.7)	102 (46.8)
Multiparous	1 (1.1)	4 (3.2)	5 (2.3)
Primi- and multiparous	27 (29.3)	81 (64.3)	108 (49.5)
Parity not specified	2 (2.2)	1 (0.8)	3 (1.4)

Note: *n* and % refer to absolute and relative frequencies, respectively.
Abbreviation: IQR, interquartile range.

Around one in five ($n=50$, 22.9%) studies in the sample did not define the onset of labor (Table 2). Only 2.2% ($n=2$) of the RCTs and 6.3% ($n=8$) of the observational studies gave a definition based on a cited reference. Around one in 10 ($n=20$, 9.2%) studies in the sample used administrative criteria, such as women's admission to the hospital, to determine the onset of labor, and the frequency was similar in both designs (RCTs: 9, 9.8% and observational studies: 11, 8.7%). Most studies that defined labor duration or its onset omitted to provide a reference: 12% ($n=11$) of the RCTs and 29.4% ($n=37$) of the observational studies.

Concerning the 79 studies investigating the active phase of the first stage of labor, 12.2% ($n=6$) of the 49 RCTs and 13.3% ($n=4$) of the 30 observational studies did not define the active phase (Table 2). Almost 80% ($n=39$) of the 49 RCTs that studied the active phase gave a definition without references, as opposed to 8.2% ($n=4$) that gave a definition with references. Among the 30 observational studies, 83.3% ($n=25$) defined it without any references, contrary to 3.3% ($n=1$) that gave a definition based on references.

In studies observing the second stage of labor ($n=158$), 5.7% ($n=4$) of the 70 RCTs defined this phase with references and 41.4% ($n=29$) without references, while 52.9% ($n=37$) gave no definition. Of the 88 observational studies that investigated the second phase, 44.3% ($n=39$) gave no definition, 46.6% ($n=41$) gave a definition without references, and 9.1% ($n=8$) gave a definition with references.

3.4 | Descriptive statistics and statistical testing

The mean was the preferred measure to describe labor duration in RCTs ($n=82$, 89.1%), followed by the standard deviation ($n=73$, 79.3%) and median ($n=20$, 21.7%) (Table 3). Contrarily, median was the preferred measure to describe labor duration in observational studies ($n=69$, 54.8%), followed by the mean ($n=64$, 50.8%) and the standard deviation ($n=56$, 44.4%). Other less popular reported measures that described the labor duration distribution included the interquartile range, range (i.e., minimum and maximum), percentile, and 95% confidence interval. The range and interquartile range were preferred in RCTs (12.0% and 8.7%, respectively), followed by the 95% confidence interval and 95th percentile (4.3% and 1.1%, respectively). On the contrary, the 95th percentile and range (24.6% and 23.0%, respectively) were reported mostly in observational studies, followed by the interquartile range and 95% confidence interval (10.3% and 4.8%, respectively). Five (2.3%) studies in the sample did not provide any measure of labor duration.

TABLE 2 Summary of the definition of labor-related concepts.

Characteristics	Randomized controlled studies ($n=92$)	Observational studies ($n=126$)	Whole sample ($n=218$)
Phase/stage analyzed *n* (%)[a]			
Whole process of labor	30 (32.6)	36 (28.6)	66 (30.3)
Latent phase	3 (3.3)	8 (6.3)	11 (5.0)
First stage	31 (33.7)	58 (46.0)	89 (40.8)
Active phase	49 (53.3)	30 (23.8)	79 (36.2)
Second stage	70 (76.1)	88 (69.8)	158 (72.5)
Onset of labor *n* (%)			
Defined with references	2 (2.2)	8 (6.3)	10 (4.6)
Defined without references	11 (12)	37 (29.4)	48 (22)
Defined with administrative criteria	9 (9.8)	11 (8.7)	20 (9.2)
Not defined	20 (21.7)	30 (23.8)	50 (22.9)
Not investigated	50 (54.3)	40 (31.8)	90 (41.3)
Active phase *n* (%)			
Defined with references	4 (4.3)	1 (0.8)	5 (2.3)
Defined without references	39 (42.4)	25 (19.8)	64 (29.4)
Not defined	6 (6.5)	4 (3.2)	10 (4.6)
Phase not investigated	43 (46.8)	96 (76.2)	139 (63.7)
Second phase *n* (%)			
Defined with references	4 (4.3)	8 (6.3)	12 (5.5)
Defined without references	29 (31.5)	41 (32.5)	70 (32.1)
Not defined	37 (40.2)	39 (31)	76 (34.9)
Phase not investigated	22 (24)	38 (30.2)	60 (27.5)

Note: *n* and % refer to absolute and relative frequencies, respectively.

[a]Some studies investigated more than one phase or stage. To obtain the percentages, each "Phase/stage analyzed" was divided by the size of the study design and the whole sample (i.e., 92, 126, and 218).

Interestingly, only nine (9.8% out of 92) RCTs and 10 (7.9% out of 126) observational studies reported both the mean and median to describe labor duration. Of the 215 studies reporting the mean and/or median, 4.3% ($n=4$) of the 92 RCTs and 26% ($n=32$) of the 123 observational studies provided a justification for the decision to report these measures of central tendency (Table 3).

The Kaplan–Meier estimator was reported in 10.9% ($n=10$) of the RCTs and 19% ($n=24$) of the observational studies (Table 3). In this sub-sample of studies ($n=34$), 50% ($n=17$) of the studies had censoring for the cesarean, followed by 38.2% ($n=13$) for instrumental deliveries and 11.8 ($n=4$) for death in childbirth.

Most studies ($n=155$, 71.1%) conducted a bivariate analysis (Table 3). In both study designs, the *t*-test was the most frequently pursued bivariate method (45.7% of RCTs and 27% of the observational studies), followed by the Mann–Whitney-Wilcoxon test (18.5% of RCTs and 19.8% of observational studies), the one-way ANOVA (9.8% of RCTs and 7.1% of observational studies) and the log-rank test (4.3% of RCTs and 6.3% of observational studies). Overall, 17.4% ($n=16$) of the RCTs and 19.8% ($n=25$) of the observational studies did not specify which bivariate method was used to compare groups (Table 3). Only 23.9% ($n=22$) of the RCTs justified the choice of the selected bivariate analysis compared with 44.4% ($n=56$) of the observational studies.

3.5 | Multivariable analysis

More than half of the observational studies applied a multivariable model ($n=66$, 52.4%) as opposed to RCTs ($n=10$, 10.9%) (Table 4). Of the 66 observational studies that used a regression model, around one in three applied interval-censored regression ($n=22$, 33.3%), followed by linear regression ($n=18$, 27.3%), Cox proportional hazards regression ($n=15$, 22.7%), and binary logistic regression ($n=11$, 16.7%). Of the 10 RCTs that used a regression model, seven applied linear regression,

TABLE 3 Summary of descriptive statistics and statistical testing.

Characteristics	RCTs ($n=92$)	OS ($n=126$)	Whole sample ($n=218$)
Measured specified n (%)[a]			
Mean	82 (89.1)	64 (50.8)	146 (67.0)
Median	20 (21.7)	69 (54.8)	89 (40.8)
Interquartile range	8 (8.7)	13 (10.3)	21 (9.6)
Range (minimum, maximum)	11 (12.0)	29 (23.0)	40 (18.3)
Standard deviation	73 (79.3)	56 (44.4)	129 (59.2)
95th percentile	1 (1.1)	31 (24.6)	32 (14.7)
95% confidence interval	4 (4.3)	6 (4.8)	10 (4.6)
Measure not specified	1 (1.1)	4 (3.2)	5 (2.3)
Choice of the mean and/or median explained n (%)			
Yes	4 (4.3)	32 (25.4)	36 (16.5)
No	88 (95.7)	94 (74.6)	182 (83.5)
Kaplan–Meier curve or estimator n (%)			
Yes	10 (10.9)	24 (19)	34 (15.6)
No	82 (89.1)	102 (81)	184 (84.4)
Which women have been right-censored n (%)			
Those with cesarean birth	5 (5.4)	12 (9.5)	17 (7.8)
Those with operative delivery	3 (3.3)	10 (7.9)	13 (6.0)
Not applicable[b]	84 (91.3)	104 (82.6)	188 (86.2)
Bivariate analyses n (%)			
t-test	42 (45.7)	34 (27)	66 (30.3)
One-way ANOVA	9 (9.8)	9 (7.1)	18 (8.3)
Mann–Whitney–Wilcoxon test	17 (18.5)	25 (19.8)	42 (19.2)
Kruskal–Wallis test	3 (3.3)	5 (4.0)	8 (3.7)
Log-rank test	4 (4.3)	8 (6.3)	12 (5.5)
Chi-squared test	1 (1.1)	3 (2.4)	4 (1.8)
Spearman's rank correlation	0 (0.0)	5 (4.0)	5 (2.3)
Not specified[c]	16 (17.4)	37 (29.4)	63 (28.8)
Choice of bivariate analysis justified n (%)			
Yes	22 (23.9)	56 (44.4)	78 (35.8)
No	54 (58.7)	45 (35.8)	99 (45.4)
Not applicable	16 (17.4)	25 (19.8)	41 (18.8)

Note: n and % refer to absolute and relative frequencies, respectively.

Abbreviations: OS, observational studies; RCT, randomized controlled studies.

[a] Some studies reported more than one measure. To obtain the percentages, each "Measured specified" was divided by the size of the study design and the whole sample (i.e., 92, 126, and 218).

[b] Studies that did not use Kaplan–Meier.

[c] Twelve observational studies did not conduct group comparisons.

and three Cox proportional hazards regression. None of the RCTs pursued binary logistic regression or interval-censored regression. Looking at the 76 (34.9%) studies that applied a multivariable regression model, all 10 RCTs and almost all observational studies ($n=62$, 94%) stratified by parity by including this covariable in the corresponding model.

4 | DISCUSSION

Like other systematic reviews, such as Hanley et al.[13] and Abalos et al.,[11] our study highlighted a lack of agreement overall with respect to how labor and its phases are defined. When such a definition was provided, most studies omitted to acknowledge the relevant source, contributing furthermore to the ongoing heterogeneity in defining labor and the phases thereof.

The present study found the median and mean labor duration to be the most preferred measures in the observational studies (54.8% and 50.8%, respectively). Abalos et al.[11] found that less than 20% of the observational studies reported the median, and more than 80% reported the mean. This disconcordance between our results and those of Abalos et al.[11] could be explained by the fact that we included more recently published studies (1999–2019) than Abalos et al.,[11] who included studies published from 1960 to 2018, with 63% of their studies being published before 2000.[11] This changing trend concerning the reporting of mean and median labor duration may be attributed to some articles on the topic, such as Vahratian et al.[12] The authors recommended the median as the appropriate statistic to describe time-to-event data with an inherent right-tail skewness, such as the labor duration.[1,11,12,16] Nevertheless, central tendency measures do not convey any information about the spread of the data. Therefore, reporting the central tendency (ideally, mean and median labor duration) with a dispersion measure (e.g., 95% percentiles) would also allow for determining physiological outliers in length.[30] Recent studies, such as that from Tilden et al.[31] or Lundborg et al.,[32] reported both mean and median alongside the dispersion measure. A thorough analysis of the right tail of the labor duration distribution is key to distinguishing between "slow but normal progress" and prolonged labor associated with poor maternal/fetal/neonatal outcomes. Tilden et al.[6] highlighted that physiological birth may still be possible beyond the 95th percentile. In their sample, the extreme right-skewed distribution persisted even in completely transfer-free and complication-free labors.

In our sample, all the studies reporting to have conducted a survival analysis method were published after 2000. Abraira et al.[33] performed a systematic survey on

TABLE 4 Summary of multivariable modeling.

Characteristics	RCTs ($n=92$)	Observational studies ($n=126$)	Whole sample ($n=218$)
Regression model used n (%)			
Linear regression	7 (7.6)	18 (14.3)	25 (11.5)
Cox regression	3 (3.3)	15 (11.9)	18 (8.3)
Binary logistic regression	0 (0.0)	11 (8.7)	11 (5.0)
Interval-censored regression	0 (0.0)	22 (17.5)	22 (10.1)
Not specified	82 (89.1)	60 (47.6)	142 (65.1)
Parity stratification n (%)			
Yes	10 (10.9)	62 (49.2)	72 (33.0)
No	0 (0.0)	4 (3.2)	4 (1.9)
Not applicable	82 (89.1)	60 (47.6)	142 (65.1)

Note: n and % refer to absolute and relative frequencies, respectively.

the reporting quality of survival analysis based on 13 high-impact medical journals. They found a changing trend in the analysis methods of the studies published in 1991 and 2007, with survival analysis techniques doubling from 17% in 1991 to 33.5% in 2007.[33] Survival analysis in medical research may have increased in the last 30 years.[33] However, our study revealed that survival analysis methods, such as the Kaplan–Meier estimator, log-rank test, Cox regression, and interval-censored regression, applied to investigate labor duration were underrepresented overall. Abraira et al.[33] found a higher proportion of multivariable analysis, with 46.2% of articles published in 1991 and 78.3% of studies published in 2007 conducting a regression analysis with almost all the studies using the Cox proportional hazard model. However, the authors considered only articles in internal medicine, cardiology, nephrology, and oncology; hence, their results may not generalize to other medical specialties.

Our study revealed that all RCTs and 94% of the observational studies with a multivariable model for labor duration controlled for parity. Controlling for confounders through a multivariable model is crucial in observational studies for lacking a randomization procedure. Nonetheless, selecting the appropriate statistical model is not a straightforward task and requires the involvement of a biostatistician in the development of the study protocol, the conduct, analysis, and reporting of the study.[34] The impact of statistical approaches on labor duration studies has currently gained interest.[35–37] He et al.[38] provided a comprehensive survey of statistical models recently used to study labor progress and highlighted the strengths and limitations of the different methods.

The limitations of the present study should be emphasized. The high number of studies conducted on labor duration during the last 20 years led us to exclude studies with labor duration as a secondary outcome. We did not evaluate the reporting quality of the studies using a suitable reporting guideline, such as the CONSORT statement for RCTs[39] and the STROBE statement for observational studies.[40] We also did not assess the internal validity of the studies, for instance, using the Cochrane risk of bias tool for the RCTs[41] and the Newcastle–Ottawa Scale for observational studies.[42] Therefore, the reporting and analysis quality of the studies may be confounded by their reporting transparency and internal validity since poorly designed and conducted studies are more likely to provide suboptimal reporting and analysis, implicating the quality of conclusions delivered to the end-users. We opted for an exploratory empirical evaluation review to assess the reporting and statistical analysis quality of the studies on labor duration published between 1999 and 2019. Then, our list with devised extraction items can be used or extended to more recent studies on labor duration to shed light on the current reporting and analysis practices.

The present study aimed to draw a picture of what has been done in our collection of studies on labor duration rather than recommend specific statistical strategies for midwifery research. Furthermore, research on integrating statistical cut-points with clinical outcomes to reach a practical definition of labor abnormalities should be initiated.[38] Since suboptimal reporting has been considered an important issue in the research of labor duration, a checklist or guidance on best reporting practices is necessary. Park et al.[43] published a list of mistakes that should be avoided for accurate and transparent reporting of survival analysis. Abraira et al.[33] also proposed a minimal set of requirements to enhance the reporting quality of survival analyses.

4.1 | Conclusions

There is an ongoing lack of agreement about the definition and diagnosis of the onset of labor and its phases. Few studies have used a survival analysis method suitable for time-to-event data. The active involvement of a biostatistician in all study stages is pivotal for quality research in labor duration. A checklist adapted to labor duration studies could be valuable to guide researchers in ensuring optimal reporting of their study reports.

ACKNOWLEDGMENTS

The authors would like to acknowledge Mrs. Cécile Jaques (Medical Library, Research and Education Department, Lausanne University Hospital, Switzerland) for the assistance in the literature search and Aurelie Abinal Delouane for her participation in the article selection process. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Mechthild M. Gross https://orcid.org/0000-0001-6348-0054

REFERENCES

1. Neal JL, Lowe NK, Ahijevych KL, Patrick TE, Cabbage LA, Corwin EJ. "Active labor" duration and dilation rates among low-risk, nulliparous women with spontaneous labor onset: a systematic review. *J Midwifery Womens Health*. 2010;55(4):308-318. doi:10.1016/j.jmwh.2009.08.004
2. National Institute for Care Excellence (NICE). Intrapartum care for healthy women and babies. Clinical guideline [CG190]. Last updated: February 21, 2017. https://www.nice.org.uk/guidance/cg190
3. Petersen A, Poetter U, Michelsen C, Gross MM. The sequence of intrapartum interventions: a descriptive approach to the cascade of interventions. *Arch Gynecol Obstet*. 2013;288(2):245-254. doi:10.1007/s00404-013-2737-8
4. Oladapo OT, Diaz V, Bonet M, et al. Cervical dilatation patterns of 'low-risk' women with spontaneous labour and normal perinatal outcomes: a systematic review. *BJOG*. 2018;125(8):944-954. doi:10.1111/1471-0528.14930
5. Gross MM, Drobnic S, Keirse MJ. Influence of fixed and time-dependent factors on duration of normal first stage labor. *Birth*. 2005;32(1):27-33. doi:10.1111/j.0730-7659.2005.00341.x
6. Tilden EL, Snowden JM, Bovbjerg ML, et al. The duration of spontaneous active and pushing phases of labour among 75,243 US women when intervention is minimal: a prospective, observational cohort study. *eClinicalMedicine*. 2022;48:101447. doi:10.1016/j.eclinm.2022.101447
7. Petersen A, Ayerle GM, Frömke C, Hecker H, Gross MM, ProGeb Study Team. The timing of interventions during labour: descriptive results of a longitudinal study. *Midwifery*. 2011;27(6):e267-e273. doi:10.1016/j.midw.2010.10.017
8. Grylka-Baeschlin S, Mueller A. Symptoms of onset of labour and early labour: a scoping review. *Women Birth*. 2023;36:483-494. doi:10.1016/j.wombi.2023.03.009
9. Grylka-Baeschlin S, Mueller AN. Primiparous women's expectations and experiences of early labour: a qualitative study. *Sex Reprod Healthc*. 2023;36:100839. doi:10.1016/j.srhc.2023.100839
10. Grylka-Baeschlin S, Gross MM, Mueller AN, Pehlke-Milde J. Development and validation of a tool for advising primiparous women during early labour: study protocol for the GebStart study. *BMJ*. 2022;12(6):e062869. doi:10.1136/bmjopen-2022-062869
11. Abalos E, Oladapo OT, Chamillard M, et al. Duration of spontaneous labour in 'low-risk' women with 'normal' perinatal outcomes: a systematic review. *Eur J Obstet Gynecol Reprod Biol*. 2018;223:123-132. doi:10.1016/j.ejogrb.2018.02.026
12. Vahratian A, Troendle JF, Siega-Riz AM, Zhang J. Methodological challenges in studying labour progression in contemporary practice. *Paediatr Perinat Epidemiol*. 2006;20(1):72-78. doi:10.1111/j.1365-3016.2006.00696.x
13. Hanley GE, Munro S, Greyson D, et al. Diagnosing onset of labor: a systematic review of definitions in the research literature. *BMC Pregnancy Childbirth*. 2016;16:71. doi:10.1186/s12884-016-0857-4
14. National Institute for Care Excellence (NICE). *Intrapartum Care: Care of Healthy Women and Their Babies during Childbirth, NICE Clinical Guidelines, No. 55*. RCOG Press; 2007. Accessed July 10, 2019. http://www.ncbi.nlm.nih.gov/pubmed/21250397
15. WHO. WHO Recommendation on Definition and Duration of the Second Stage of Labour |RHL. 2018. WHeO Reproductive Health Library. https://extranet.who.int/rhl/topics/preconception-pregnancy-childbirth-and-postpartum-care/care-during-childbirth/care-during-labour-2nd-stage/who-recommendation-definition-and-duration-second-stage-labour
16. Zhang J, Troendle JF, Yancey MK. Reassessing the labor curve in nulliparous women. *Am J Obstet Gynecol*. 2002;187(4):824-828. doi:10.1067/mob.2002.127142
17. Friedman EA. An objective approach to the diagnosis and management of abnormal labor. *Bull N Y Acad Med*. 1972;48(6):842-858.
18. Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part I: basic concepts and first analyses. *Br J Cancer*. 2003;89(2):232-238. doi:10.1038/sj.bjc.6601118
19. Goel MK, Khanna P, Kishore J. Understanding survival analysis: Kaplan-Meier estimate. *Int J Ayurveda Res*. 2010;1(4):274-278. doi:10.4103/0974-7788.76794
20. Singh R, Mukhopadhyay K. Survival analysis in clinical trials: basics and must know areas. *Perspect Clin Res*. 2011;2(4):145-148. doi:10.4103/2229-3485.86872
21. Zhang J, Landy HJ, Ware Branch D, et al. Contemporary patterns of spontaneous labor with normal neonatal outcomes. *Obstet Gynecol*. 2010;116(6):1281-1287. doi:10.1097/AOG.0b013e3181fdef6e
22. Bland JM, Altman DG. The logrank test. *BMJ*. 2004;328(7447):1073. doi:10.1136/bmj.328.7447.1073
23. Rutherford MJ, Lambert PC, Sweeting MJ, et al. NICE DSU Technical Support Document 21. Flexible Methods for Survival Analysis. 2020. http://www.nicedsu.org.uk
24. Gross MM, Frömke C, Hecker H. The timing of amniotomy, oxytocin and neuraxial analgesia and its association with

labour duration and mode of birth. *Arch Gynecol Obstet*. 2014;289(1):41-48. doi:10.1007/s00404-013-2916-7
25. Grylka-Baeschlin S. Timing of interventions and events associated with labour duration and mode of birth in women with planned vaginal births after caesarean section. PhD Thesis – Recherche Google. Hannover Medical School. 2017. Accessed July 6, 2020. https://www.google.com/search?client=safari&rls=en&q=Timing+of+interventions+and+events+associated+with+labour+duration+and+mode+of+birth+in+women+with+planned+vaginal+births+after+caesarean+section.+PhD+Thesis&ie=UTF-8&oe=UTF-8
26. Bradburn MJ, Clark TG, Love SB, Altman DG. Survival analysis part II: multivariate data analysis–an introduction to concepts and methods. *Br J Cancer*. 2003;89(3):431-436. doi:10.1038/sj.bjc.6601119
27. Prinja S, Gupta N, Verma R. Censoring in clinical trials: review of survival analysis techniques. *Indian J Community Med*. 2010;35(2):217-221. doi:10.4103/0970-0218.66859
28. Bradburn MJ, Clark TG, Love SB, Altman DG. Survival analysis Part III: multivariate data analysis – choosing a model and assessing its adequacy and fit. *Br J Cancer*. 2003;89(4):605-611. doi:10.1038/sj.bjc.6601120
29. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol*. 2009;62(10):1006-1012. doi:10.1016/j.jclinepi.2009.06.005
30. Lee AH, Fung WK, Fu B. Analyzing hospital length of stay: mean or median regression? *Med Care*. 2003;41(5):681-686. doi:10.1097/01.MLR.0000062550.23101.6F
31. Tilden EL, Caughey AB, Ahlberg M, et al. Latent phase duration and associated outcomes: a contemporary, population-based observational study. *Am J Obstet Gynecol*. 2023;228(5S):S1025-S1036.e9. doi:10.1016/j.ajog.2022.10.003
32. Lundborg L, Åberg K, Sandström A, et al. First stage progression in women with spontaneous onset of labor: a large population-based cohort study. *PLoS One*. 2020;15(9):e0239724. doi:10.1371/journal.pone.0239724
33. Abraira V, Muriel A, Emparanza JI, et al. Reporting quality of survival analyses in medical journals still needs improvement. A minimal requirements proposal. *J Clin Epidemiol*. 2013;66(12):1340-1346.e5. doi:10.1016/j.jclinepi.2013.06.009
34. Zapf A, Rauch G, Kieser M. Why do you need a biostatistician? *BMC Med Res Methodol*. 2020;20(1):23. doi:10.1186/s12874-020-0916-4
35. de Vries BS, Mcdonald S, Joseph FA, et al. Impact of analysis technique on our understanding of the natural history of labour: a simulation study. *BJOG*. 2021;128(11):1833-1842. doi:10.1111/1471-0528.16719
36. Zhang J, Troendle J, Souza JP, Oladapo OT. Re: impact of analysis technique on our understanding of the natural history of labour. *BJOG*. 2022;129(11):1939-1940. doi:10.1111/1471-0528.17221
37. Hamilton EF, Romero R, Tarca AL, Warrick PA. The evolution of the labor curve and its implications for clinical practice: the relationship between cervical dilation, station, and time during labor. *Am J Obstet Gynecol*. 2023;228(5S):S1050-S1062. doi:10.1016/j.ajog.2022.12.005
38. He X, Zeng X, Troendle J, et al. New insights on labor progression: a systematic review. *Am J Obstet Gynecol*. 2023;228(5S):S1063-S1094. doi:10.1016/j.ajog.2022.11.1299
39. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med*. 2010;152(11):726-732. doi:10.7326/0003-4819-152-11-201006010-00232
40. von Elm E, Altman DG, Egger M, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med*. 2007;147(8):573-577.
41. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. 2019;366:l4898.
42. Wells G, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2013. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp
43. Park SH, Han K, Park SY. Mistakes to avoid for accurate and transparent reporting of survival analysis in imaging research. *Korean J Radiol*. 2021;22(10):1587-1593. doi:10.3348/kjr.2021.0579

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Celetta E, Spineli LM, Avignon V, Gehling H, Gross MM. An exploratory review on the empirical evaluation of the quality of reporting and analyzing labor duration. *Birth*. 2024;51:773-782. doi:10.1111/birt.12833