ESTIMATING THE PROPORTION OF CLINICAL MASTITIS ATTRIBUTABLE TO SUBCLINICAL MASTITIS IN DAIRY CATTLE USING TWO MULTIVARIABLE STATISTICAL APPROACHES

B.H.P. VAN DEN BORNE*, J.C.M. VERNOOIJ, A.M. LUPINDU, G. VAN SCHAIK, K. FRANKENA, T.J.G.M. LAM AND M. NIELEN

SUMMARY

The aim of this study was to quantify the proportion of first bovine CM cases attributable to high composite somatic cell counts (CSCC). Cows were followed from the first CSCC measurement postpartum until CM or censoring, using survival analysis. A conditional logistic regression model was also fitted to the data with CM cows being matched to cows without CM. Both models identified high CSCC cows to have a higher risk for subsequent CM cases than low CSCC cows. The population attributable fraction was 0.22 for primiparae and 0.17 for multiparae according to the logistic regression model, while it was 0.25 in the survival analysis. The latter approach also identified that the proportion of cows without CM would increase from 89% to 93%. Both multivariable statistical approaches showed that a substantial reduction in CM can be achieved by decreasing the prevalence of high CSCC in the dairy population.

INTRODUCTION

Mastitis, the inflammation of the mammary gland, generally follows infection. When inflammation is accompanied with visible alterations of the udder and/or milk it is called clinical mastitis (CM). Subclinical mastitis is defined as a mammary gland that is inflamed, but has no visible signs. In dairy cows, both appearances are associated with economic losses (e.g., milk yield loss, antimicrobial use, culling, extra labour; Halasa et al., 2007). Bovine subclinical mastitis is generally monitored by measuring the composite somatic cell count (CSCC) during the regular test day recording (Schukken et al., 2003). High CSCC are measured before CM occurs (de Haas et al., 2002) and are therefore considered predictive for the development of CM (e.g., Rupp and Boichard, 2000; Green et al., 2004). Several statistical methods have previously been used to determine this relationship, but most investigations neither corrected for the length of the time period up to CM occurrence nor adjusted for the dynamic nature of CSCC. Intramammary infections can occur during lactation resulting in new CSCC elevations (Schukken et al., 2003).

*Bart van den Borne, Faculty of Veterinary Medicine, Utrecht University, Yalelaan 7, 3584 CL Utrecht, the Netherlands. Currently at: Veterinary Public Health Institute, Vetsuisse Faculty, University of Bern, Schwarzenburgstrasse 155, 3097 Liebefeld, Switzerland. Email: bart.borne@vphi.unibe.ch

The population attributable fraction (PAF) is the proportion of disease events that can be prevented from the total population when a perfect intervention is applied to a certain risk factor (Dohoo et al., 2003). PAF estimation is important to determine the (economic) impact of intervention strategies on population level because it, in contrast to the attributable fraction, it includes the exposed as well as the non-exposed individuals. Several techniques have been proposed to calculate PAF but only some of them adjust for other risk factors in case of multivariable modelling (Rückinger et al., 2009). Moreover, the PAF becomes dynamic with censored time-to-event data as the event rate accumulates over time and exposed subjects are more rapidly depleted from the population than non-exposed subjects (Cox et al., 2009). For the purpose of high CSCC in dairy cows, PAF indicates the potential reduction in CM events in the total dairy cow population, by preventing or removing high CSCC using intervention (e.g., culling or antimicrobial treatment). The objective of this study was to estimate the proportion of CM cases attributable to high CSCC in Dutch dairy herds using two statistical multivariable approaches.

MATERIALS AND METHODS

Data

The data for this investigation were obtained from a previously described observational study on 205 randomly selected Dutch dairy herds participating in four-weekly test recording (van den Borne et al., 2010). In short, CSCC measurements of the regular test day recording and farmer-diagnosed CM cases on these herds were collected from July 1, 2004 until June 30, 2005. Test day recordings and CM cases were selected from cows that calved within the study period to evaluate the association between CSCC and the first subsequent case of CM. Recurrent CM cases were not included in the analysis, nor were CM cases before the first test day. Cows without a test day record preceding a CM case (n=892) were removed from the dataset. The first lactation was analysed when cows had two calvings within the study period (n=48 cows). This dataset comprised cow level information on stage of lactation, parity, CSCC and production parameters at each test day. Seventeen mastitis management related factors on herd level (including herd size, bulk milk SCC measurements, and factors on hygienic procedures and treatment decisions) were obtained from a questionnaire sent to the farmers before the start of the study (Jansen et al., 2009) to correct for the influence of these herd level factors on the relationship between CSCC and CM. Nine herds were removed from the data because of missing herd level data. This resulted in a dataset of 13,917 cows in 196 herds, of which 1,560 cows (11.2%) had a first case of CM after a CSCC measurement. This was the initial dataset for both analytical procedures.

Survival analysis

Cox proportional hazard regression was used to determine how the hazards for CM differed for cows with low (<200,000 cells/ml) or high (≥200,000 cells/ml) CSCC. CSCC status and other test day parameters (number of test days since calving (used as a proxy for lactation stage); milk production (kg/day); protein percentage; fat percentage; and season (pasture period: May-October or housing period: November-April)) were included as time-varying predictors that could change at each test day. Each cow became at risk at the first test day after calving and its failure time was determined in each test day interval until the first case of CM or until censoring at the end of the study period (July 1, 2005) or at the last available test day if cows were culled or dried off. All available risk factors (6 time-varying predictors, parity, and 17 herd level risk factors) were tested in bivariable analyses with CSCC status forced into the model. The

reduction in deviance was used to select variables with $P < 0.25$ for the multivariable analyses, in which a stepwise backward elimination process was used to identify the variables that were significantly ($P < 0.05$) contributing to the model. A herd frailty effect was added to all multivariable models to adjust for clustering within herds. All models were checked for confounding, which was assumed to occur when estimates changed >25%. Biologically relevant interactions between CSCC status and other cow level factors were also investigated. Survival analysis was performed in R (R Development Core Team, 2009).

Conditional logistic regression

In addition to the Cox proportional hazards model, a matched case-control study design was also applied to the initial dataset in which a case was defined to be a cow with a first CM case after at least 1 CSCC measurement. CSCC status of these CM cases was determined at the last test day preceding the first CM observation (td = -1). The preceding test day was evaluated if CM occurred on a test day. A cow was considered to have high CSCC when CSCC was ≥200,000 cells/ml, similar to the definition used for the survival analysis. Test day recordings from control cows in the same stage of lactation, without CM in the study, were matched to td = -1 from case cows. Therefore, eligible control cows needed to have their calving dates within 7 days around the calving date of a case cow. Additionally, the test day recordings from control cows needed to be within 7 days of td = -1 of their case cow and were subsequently also defined td = -1. CSCC status at td = -1 for control cows was determined similar to case cows. This matching procedure adjusted for potential differences in CSCC between lactation stages. Four control cows from any herd were randomly selected to each case cow to increase precision. Control cows were matched to only 1 case cow.

All available risk factors were studied using conditional logistic regression with CM occurrence as the binary response variable. Lactation stage was included in the analysis to verify if matching did not result in confounding with the risk factor exposure (i.e., CSCC status). Analyses started with univariable and bivariable conditional logistic regression models for all eligible variables. CSCC status was forced into all models to observe the effect of the variable of interest on the parameter estimate of CSCC status. All variables with a $P < 0.25$ qualified for the multivariable analysis, in which a stepwise backwards approach was used until all variables had a $P < 0.05$ based on the reduction in deviance. Observations with missing values for some herd level variables were added to the dataset again when these variables were not included in the model anymore. All models were checked for confounding, which was assumed to occur when estimates changed >25%. Biologically relevant interactions between cow level variables were also tested. No random herd effect could be added to the model due to the matched case-control study design and the accompanying analytical procedure (i.e., conditional logistic regression). Standard errors of herd level variables in the final model may therefore be underestimated resulting in too low P values. Consequently, the herd level variables in the final model were not interpreted or further tested in interaction terms, but were assumed to adjust the effect estimates of CSCC status for a potential herd effect. Conditional logistic regression analysis was performed using proc logistic in SAS 9.2 (SAS Institute, Cary, USA).

Population attributable fraction

For the Cox frailty model, the PAF was calculated as follows (Cox et al., 2009):

$$PAF(t) = \frac{S^*(t) - S(t)}{1 - S(t)}$$

with $S(t) = \sum_j \rho_j S_j(t)$ the survival function of the total population (both exposed and unexposed individuals) for j strata, $S^*(t) = \sum_j \rho_j^* S_j(t)$ the resulting survival distribution for j strata when a perfect intervention of the exposure variable (i.e., high CSCC) is assumed to result in an alternative distribution of ρ_j^* (absence of high CSCC), and ρ_j the proportion of exposed individuals in stratum *j*.

PAF(t) indicates the reduction in diseased individuals, but interest mainly is in survival of individuals when conducting survival analysis. Cox et al. (2009) therefore introduced a new measure of association to assess the impact of interventions with survival data: the attributable survival (hereafter called the population attributable survival (PAS), in agreement with PAF). PAS represents the additional proportion of individuals in the population who survive to a given time, if a fully effective intervention to exposed individuals has been administered at t = 0. For the current study, PAS represents the additional proportion of cows without subsequent CM after a CSCC measurement when preventing or removing high CSCC from the population. PAS is calculated as follows (Cox et al., 2009):

$$PAS(t) = \frac{S^*(t) - S(t)}{S^*(t)}$$

Both PAF(t) and PAS(t) provide information about the timing of intervention because the survival function is dependent on t (Cox et al., 2009). PAF(t) en PAS(t) calculations were based on the final Cox proportional hazards model without the herd level frailty effect included.

The 'average' PAF for high CSCC was calculated according to Eide and Gefeller (1995) in the unconditional logistic regression analysis, using a readily available SAS macro (Rückinger et al., 2009). Only the cow level variables and the interaction between CSCC status and parity were included in the PAF calculations because of computational capacity. The change in PAF estimate for CSCC status, however, was determined when each herd variable was added one at the time to the model.

RESULTS

Survival analysis

Survival without CM differed between parities. Survival without CM was the highest in primi- and multiparae with a low CSCC, while survival without CM was the lowest in multiparae with a high CSCC (Fig. 1).

The results from the final Cox proportional hazards model are presented in Table 1 and are based on 62,742 test day intervals, of which 1,280 (2.0%) had a first case of CM. High CSCC cows had a 2 to 4 fold higher hazard for subsequent CM than cows with a low CSCC. Multiparae with a low CSCC had a 2 fold higher hazard for CM than primiparae with a low CSCC. Higher production levels were associated with increased hazards for CM, especially in high CSCC cows. The hazard in the housing period was slightly higher than in the pasture period. No herd level variables were significant in the final Cox frailty model. Variance of the herd frailty effect was 0.37 ($P < 0.0001$).

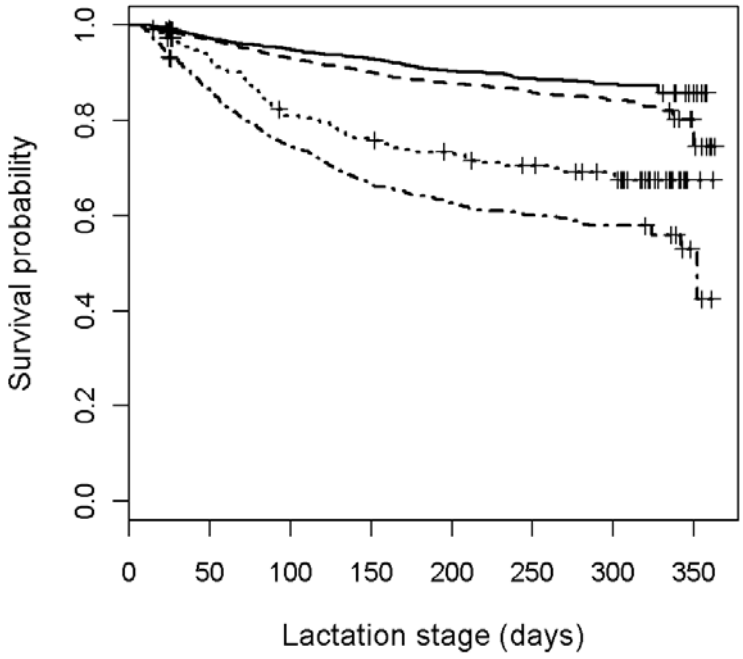

Fig. 1 Survival curves (not having clinical mastitis) for 4 groups of dairy cows, after the first somatic cell count (CSCC) measurement postpartum (from top to bottom: primiparae with a low CSCC (<200,000 cells/ml); multiparae with a low CSCC; primiparae with a high CSCC; multiparae with a high CSCC).

Table 1. Results from the final Cox proportional hazards model with time-varying predictors for the first case of clinical mastitis (CM; n=1,280) after a composite somatic cell count measurement (CSCC).

VARIABLE	CATEGORY	HR	95% - CI
High CSCC (≥200,000 cells/ml)	No	1.0	Ref
In primiparae	Yes	4.0	2.5 – 6.5
In multiparae	Yes	2.1	1.3 – 3.6
Parity	1	1.0	Ref
For low CSCC cows	≥2	1.9	1.6 – 2.4
For high CSCC cows	≥2	1.0	0.8 – 1.3
Season	Pasture	1.0	Ref
	Housing	1.2	1.0 – 1.3
Milk production (kg/day)	Continuous		
In low CSCC cows		1.02	1.00 – 1.03
In high CSCC cows		1.04	1.03 – 1.05
Herd frailty effect			

HR = Hazard Ratio; 95% - CI = 95% - confidence interval. Ref = Reference category.

The PAF(t) and PAS(t) based on the final Cox proportional hazards model are presented in Fig. 2. The PAF starts at 0.29 at the beginning of the lactation and is monotone decreasing to 0.25 towards 300 days in lactation. The PAS starts at zero directly after calving and is monotone increasing towards 0.04 at 300 days in lactation. This indicates that an additional 4% of cows do not develop a first subsequent CM after a high CSCC measurement when the latter can be prevented. Hence, it results in a reduction of cows with a first case of CM during their lactation from 11% to 7%, thereby increasing the proportion of cows without CM from 89% to 93%.

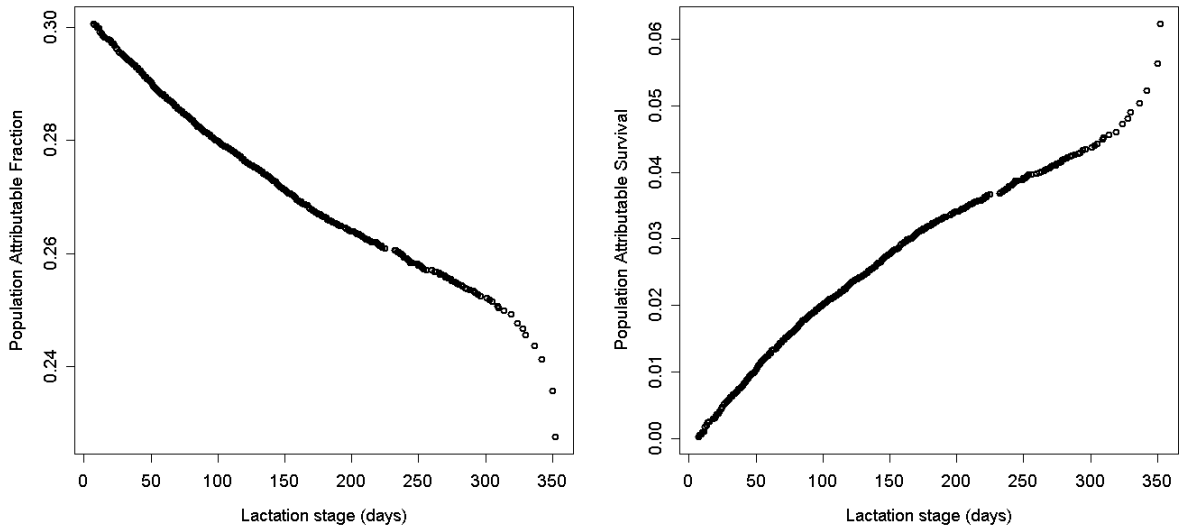

Fig. 2 Estimated population attributable fraction (left) and population attributable survival (right) for clinical mastitis in dairy cows with a high composite somatic cell count.

Conditional logistic regression

The data analyzed in the conditional logistic regression consisted of 1,560 CM cows and 6,240 selected control cows. There were no herd mates within 1,479 matched case and control sets, while there was 1 herd mate among the controls in 81 matched sets. This indicates that only a small proportion (5.2%) in matched sets came from the same herd. High CSCC at td = -1 was observed in 41.3% of the CM cows while high CSCC were observed in 15.3% of the control cows.

Results from the final conditional logistic regression model are presented in Table 2 and are based on 7,684 observations (including 1,538 cases) due to missing values for some herd level variables. All cow level variables significant in the final Cox proportional hazards model were also significantly associated with CM in the conditional logistic regression. Primi- and multiparae with a high CSCC at td = -1 had a higher odds (OR = 6 and OR = 4 for primi- and multiparae, respectively) for CM occurrence than cows with a low CSCC. CM was found more frequently in multiparae than in primiparae when they had a low CSCC (OR = 2). Crude proportions of CM occurrence were 35.5% and 42.7% in primi- and multiparae with a high CSCC, respectively, and were 8.5% and 18.3% for primi- and multiparae with a low CSCC. A linear increase in milk production was associated with a log linear increase in CM occurrence. In contrast to the Cox proportional hazards model, CM occurred slightly more frequent in the pasture period than in the housing period. Four herd level variables were significant in the final conditional logistic regression model. The number of test day recordings since calving, as a

proxy for stage of lactation, was not associated with CM occurrence, indicating proper matching of case and control cows.

Table 2. Results from the final conditional logistic regression model for the first case of clinical mastitis (CM; n=1,538) after a composite somatic cell count (CSCC) measurement.

VARIABLE	CATEGORY	OR	95% - CI
High CSCC (≥200,000 cells/ml)	No	1.0	Ref
In primiparae	Yes	6.1	4.5 – 8.3
In multiparae	Yes	3.6	3.1 – 4.2
Parity	1	1.0	Ref
For low CSCC cows	≥2	1.9	1.5 – 2.3
For high CSCC cows	≥2	1.1	0.8 – 1.5
Season	Pasture	1.0	Ref
	Housing	0.6	0.5 – 0.9
Milk production (kg/day)	Continuous	1.03	1.02 – 1.04
Average milking herd size		NI	
Post-milking teat disinfection		NI	
Treatment of cows with a 1st high CSCC		NI	
Performing a dynamic milking test		NI	

OR = Odds Ratio; 95% - CI = 95% - confidence interval. Ref = Reference category. NI = Not interpreted. Herd level variables were not interpreted because they were assumed to correct the cow level estimates for clustering within herds.

According to the logistic regression model, the 'average' PAF for high CSCC was 0.221 and 0.166 in primi- and multiparae, respectively. The maximum change in 'average' PAF for CSCC status was 5.3% when herd level variables were added to the model one at the time, indicating a robust 'average' PAF estimate.

DISCUSSION

This study quantified the relationship between high CSCC and the first subsequent CM case using Cox proportional hazards models and conditional logistic regression models. This relationship was only based on a statistical relation, while the relationship between subclinical mastitis and CM should ideally be based on the presence of bacteria in both types of mastitis. High CSCC are a good approximation for the bacteriological status of a cow's udder but its sensitivity and specificity are not perfect (Schukken et al., 2003). Nevertheless, both approaches identified high CSCC cows to have a higher risk for developing CM than cows with a low CSCC, as did other studies (Rupp and Boichard, 2000; Green et al., 2004). The identified risk factors (parity, milk yield and season) behaved as determined earlier (Rupp and Boichard, 2000; Green et al., 2004). The opposite effect of season in the conditional logistic regression model compared with the Cox proportional hazards model was probably due to the matching procedure. Case and control cows were in the same season in 85% of the matched groups, indicating limited variation within matching groups because of overmatching (case and control cows calved within 7 days of each other).

Both statistical approaches were based on the same initial dataset but the effects estimates and the PAF estimates from the Cox proportional hazards model cannot be fully compared with

the estimates from the conditional logistic regression model. There are methodological differences between both statistical approaches (e.g., matching vs no matching, time-fixed vs time-vaying predictors, no corrections for clustering within herds vs random frailty effects). Effect estimates, and thus the PAF estimates, from Cox proportional hazards models are more accurate and precise compared with effect estimates from conditional logistic regression models (Green and Symons, 1983) and are therefore preferred.

Calculation of the 'average PAF' identified that a substantial reduction of 22.1% and 16.7% of first CM cases of primi- and multiparae may be achieved by preventing, treating and/or culling of high CSCC. Other approaches to PAF calculations have been suggested previously and were reviewed recently by Rückinger et al. (2009). The 'average' PAF according to Eide and Gefeller (1995) was acknowledged to give the most plausible PAF estimates while adjusting for other risk factors. Less complicated formulas may overestimate the potential population effect (Rückinger et al., 2009). Those were therefore not applied in the current study. It has to be noted, however, that despite its superiority to other formulas, PAF interpretation is only possible for dichotomous risk factors. PAF estimates can be adjusted with non-dichotomous covariates, which were the case in the current logistic regression model that included milk production. Communicating PAF results for continuous covariates, however, may be difficult (Ruckinger et al., 2009).

Using the recently proposed formulas for PAF and PAS estimation for time-to-event data (Cox et al., 2009), PAF was 25% and PAS was 4% in the Cox proportional hazards model. This confirms the potential gain in CM in the total population when high CSCC is prevented or intervened. Both formulas assume a perfect intervention at the time of detection (i.e., td=-1 in this study), while interventions on high CSCC are neither directly applied nor perfect in dairy practice. These aspects can be taken into account using more advanced formulas (Cox et al., 2009).

In conclusion, this study quantified the relationship between high CSCC and first cases of subsequent CM. It was shown that approximately 25% of first subsequent cases of CM in dairy cows can potentially be reduced by preventing, treating or culling of high CSCC cows. This population effect is considered a substantial reduction, especially when it is considered that only first subsequent CM cases were taken into account. Cows may develop repeated cases of CM, resulting in an underestimation of the true population effect.

ACKNOWLEDGEMENTS

This study is part of the 5-year mastitis control program of the Dutch Udder Health Centre and was financially supported by the Dutch Dairy Board.

REFERENCES

Cox, C., Chu, H. and Muñoz, A. (2009). Survival attributable to an exposure. Stat. Med. 28, 3276-3293

de Haas, Y. Barkema, H.W. and Veerkamp, R.F. (2002). The effect of pathogen-specific clinical mastitis on the lactation curve for somatic cell count. J. Dairy Sci. 85, 1314-1323

Dohoo, I.R., Martin, S.W. and Stryhn, H. (2003). Veterinary Epidemiologic Research. Atlantic Veterinary College Inc., Charlottetown, Prince Edward Island, Canada

Eide, G.E. and Gefeller, O. (1995). Sequential and average attributable fractions as aids in the selection of preventive strategies. J. Clin. Epidemiol. 48, 645-655

Green, M.J., Burton, P.R., Green, L.E., Schukken, Y.H., Bradley, A.J., Peeler, E.J. and Medley, G.F. (2004). The use of Markov chain Monte Carlo for analysis of correlated binary data: patterns of somatic cells in milk and the risk of clinical mastitis in dairy cows. Prev. Vet. Med. 64, 157-174

Green, M.S. and Symons, M.J. (1983). A comparison of the logistic risk function and the proportional hazards model in prospective epidemiologic studies. J. Chronic Dis. 36, 715-724

Halasa, T., Huijps, K., Østerås, O. and Hogeveen, H. (2007). Economic effects of bovine mastitis and mastitis management: A review. Vet. Q. 29, 18-31

Jansen, J., van den Borne, B.H.P., Renes, R.J., van Schaik, G., Lam, T.J.G.M. and Leeuwis, C. (2009). Explaining mastitis incidence in Dutch dairy farming: The influence of farmers' attitudes and behaviour. Prev. Vet. Med. 92, 210-223

Rückinger, S., von Kries, R. and Toschke, A.M. (2009). An illustration of and programs estimating attributable fractions in large scale surveys considering multiple risk factors. BMC Med. Res. Methodol. 9, 7

Rupp, R. and Boichard, D. (2000). Relationship of early first lactation somatic cell count with risk of subsequent first clinical mastitis. Livest. Prod. Sci. 62, 169-180

Schukken, Y.H., Wilson, D.J., Welcome, F., Garrison-Tikofsky, L. and Gonzalez, R.N. (2003). Monitoring udder health and milk quality using somatic cell counts. Vet. Res. 34, 579-596

van den Borne, B.H.P., van Schaik, G., Lam, T.J.G.M. and Nielen, M. (2010). Variation in herd level mastitis indicators between primi- and multiparae in Dutch dairy herds. Prev. Vet. Med. 96, 49-55