Field evaluation of diagnostic sensitivity (DSe) and specificity (DSp) of common tests for amoebic gill disease (AGD) and complex gill disease (CGD) in cultured Atlantic salmon (Salmo salar) in Scotland using Bayesian latent class models

Boerlage, AS; Ashby, Angela; Eze, JI; Gunn, GJ; Reeves, A

Published in:
Preventive Veterinary Medicine

DOI:
10.1016/j.prevetmed.2022.105654

Print publication: 01/07/2022

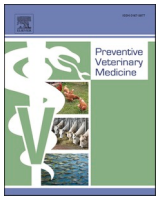
Document Version
Publisher's PDF, also known as Version of record

Preventive Veterinary Medicine 204 (2022) 105654

Contents lists available at ScienceDirect

Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed



Field evaluation of diagnostic sensitivity (DSe) and specificity (DSp) of common tests for amoebic gill disease (AGD) and complex gill disease (CGD) in cultured Atlantic salmon (*Salmo salar*) in Scotland using Bayesian latent class models

Annette S. Boerlage [a,*], Angela Ashby [b], Jude I. Eze [a,c], George J. Gunn [a], Aaron Reeves [a]

[a] *Epidemiology Research Unit, Department of Veterinary and Animal Science, Northern Faculty, Scotland's Rural College (SRUC), An Lòchran, 10 Inverness Campus, Inverness, UK*
[b] *Pharmaq Analytiq, 22 Carsegate Road, Inverness, Scotland, UK*
[c] *Biomathematics and Statistics Scotland, JCMB, The King's Buildings, Peter Guthrie Tait Road, Edinburgh EH9 3FD, UK*

ARTICLE INFO

Keywords:
Gross gill score
Histopathology
Marine gill diseases
Sensitivity
Specificity
Latent class models

ABSTRACT

Amoebic gill disease (AGD) and complex gill disease (CGD) are the most significant marine gill diseases in salmon aquaculture in Scotland. Little is published about diagnostic performance of tests to detect these diseases, making it difficult to interpret test results. We estimated diagnostic sensitivity (DSe) and specificity (DSp) of common tests for AGD (gross AGD score, qPCR for *Neoparamoeba perurans*, histopathology) and CGD (gross proliferative gill disease (PGD) score, gross total gill score, histopathology). Because specifications in our sampling protocol implemented to encourage consistency across the farms might affect diagnostic performance of histopathology (historically the reference standard for gill diseases), we used Bayesian latent class models without reference standard. Cases and non-cases were based on less, medium, and severe stringent case definitions, representing different cut-off levels for the different tests. Gross gill scores for both diseases were excellent in designating non-diseased fish, DSps were generally around 1. To detect CGD, DSe of gross total gill score and gross PGD score were between respectively 0.81 (0.73 – 0.91 lower to upper 95% credible interval) and 0.53 (0.46 – 0.64) for medium stringent case definitions, and to detect AGD the DSe for the gross AGD score was between 0.53 (0.48–0.57) and 0.14 (0.07 – 0.22) for respectively the less and severe stringent case definition. Thus, gross gill scores were medium to good in designating truly diseased fish, implying some false negatives are expected. For CGD the DSe for gross total gill scores were the highest, for AGD it was the qPCR test at a DSe of 0.92 (0.86 – 0.99). For both diseases, DSe was lowest for histopathology, e.g. 0.23 (0.16 – 0.30) for AGD and 0.1 (0.07 – 0.14) for CGD under medium stringent case definitions, perhaps due to collecting the second gill arch on the right rather than the worst affected arch, whilst PCR sampling and gross gill scoring included multiple (PCR) or all (gross scoring) gill arches. The diagnostic goals of these tests differ; gross gill scoring provides a low-cost presumptive diagnosis, PCR a non-lethal confirmation of the presence of a specific pathogen and histopathology provides information on the underlying aetiology of gill damage as well as the extent, severity, and chronology of gill disease. An effective gill health surveillance strategy is likely to incorporate multiple diagnostic tools used in a complementary manner.

1. Introduction

In the last decade, marine gill diseases in European cultured salmon have emerged and established as common recurring challenges with a significant impact (Hjeltnes et al., 2017; Matthews et al., 2013). There are currently seven distinguishable types: (i) amoebic gill disease (AGD), (ii) parasitic gill disease, (iii) viral gill disease, (iv) bacterial gill disease, (v) zooplankton (cnidarian nematocyst)-associated gill disease, (vi) harmful algal gill disease and (vii) chemical/toxin-associated gill disease (Rodger, 2007). In addition to disease caused by a single specific aetiological agent, complex gill disease (CGD) has emerged as a health concern in farmed Atlantic salmon in recent years. CGD refers to a

* Corresponding author.
E-mail address: Annette.Boerlage@sruc.ac.uk (A.S. Boerlage).

https://doi.org/10.1016/j.prevetmed.2022.105654
Received 9 November 2021; Received in revised form 15 March 2022; Accepted 18 April 2022
Available online 22 April 2022

syndrome arising when some or all the seven marine gill disease types are observed simultaneously. There may be no obvious primary causal agent, and principal pathological changes are non-specific and often proliferative (Boerlage et al., 2020). AGD and CGD are arguably the most significant marine gill diseases in Atlantic salmon aquaculture.

Amoebic gill disease is distinguishable and caused by the amoeba *Neoparamoeba perurans*. AGD is the only marine gill disease for which mitigation is a common feature on most Scottish salmon farms due to the persistence of the amoeba in the environment and effectiveness of mitigation measures such as freshwater and hydrogen peroxide baths (Rodger, 2013). Different diagnostic methods are used to test for presence and severity of AGD. Histopathology is typically considered the reference standard (Adams et al., 2004). Lesions observed in AGD cases include lamellar hyperplasia, lamellar fusion, inflammation, and presence of the amoebae (Adams and Nowak, 2001). Gross gill scoring is a test that relies on visual inspection of the gills of sedated salmon. For AGD, this include multifocal pale lesions on the gill surface and raised white mucoid spots and plaques (Adams and Nowak, 2004). Gross lesions for AGD are mostly distinctive from those of other marine gill disease types, though some pathology that can be observed in the gills during AGD gill scoring may be caused in another way (Adams et al., 2004; Mo et al., 2015). AGD lesions can become less distinctive later in the production cycle or in the presence of concurrent gill disease. The most used gross gill score protocol runs from zero to five and is developed by (Taylor et al., 2009b). Other tests for AGD are based on immunofluorescence antibody test (IFAT; not commonly used in Scotland) and PCR which indicate presence of the causative agent, *N. perurans* (Downes et al., 2015; Hellebø et al., 2017; Rozas et al., 2011).

CGD is a term which has emerged in the last decade (Herrero et al., 2018). The complex component reflects our level of understanding of the significance and interaction of putative pathogens and environmental insults (Boerlage et al., 2020; Mitchell and Rodger, 2011; Rodger et al., 2011). There currently is no specific treatment for CGD. When AGD is part of CGD, which is often the case in Scotland (Boerlage et al., 2020), treatments for AGD may affect overall CGD as well. Additional mitigation may be achieved through general adjustments in management practices known to relieve, reduce or prevent stress, such as reducing feeding levels, splitting down groups to reduce biomass density, or delaying handling events. When phytoplankton is involved, measures such as bubble curtains or additional oxygen supply to the pen may be employed (Mardones et al., 2021; Rodger et al., 2011). As for AGD, histopathology is considered the reference standard for CGD. Since gill changes are non-specific and multi factorial, a total histopathology score can be used. An example is the total histopathology score developed by (Mitchell et al., 2012a) which includes index and ancillary criteria that are combined to form a score that ranges from zero to 24. This score can include a variety of abnormalities to arrive at a similar level of severity. Another way to test for CGD is by gross gill observations. Few protocols on gill scores for CGD are published, an example is the non-specific gill score which looks at different levels of necrotic tissue in a scale from zero to five (Bloecher et al., 2018), and the proliferative gill disease (PGD) score (Król et al., 2020). PCR tests for putative pathogens have been used, but because the pathogens they target are putative and not always present in all forms of CGD, their usefulness may be limited.

The unknown diagnostic performance of the gross gill scores lead to uncertainty about their significance. Understanding the ability of a test to provide a correct designation of disease is essential for, for example, decision making around when to treat, whether a handling event would be harmful for the fish, understanding the meaning of observations in a research project, or estimating true prevalence instead of apparent prevalence (Taylor et al., 2009b). The two key characteristics are diagnostic sensitivity (DSe), the ability that a truly diseased animal will indeed be tested positive, and diagnostic specificity (DSp), the ability that a truly non-diseased animal will indeed be tested negative (Dohoo et al., 2009). A lack of DSe leads to false negatives, a lack of DSp to false positives. Historically, DSe and DSp estimations make use of a gold standard, also (and in this study) referred as reference standard. A reference standard is a test of which the outcomes are considered absolute and always true. Using histopathology as reference standard, DSe and DSp of gross AGD scores have been estimated at 0.74–0.78 and 0.75–0.82 respectively (Adams et al., 2004; Rozas et al., 2011) and for PCR 0.94 and 0.97 respectively (Rozas et al., 2011). There are to our knowledge no estimates for DSe and DSp for tests for CGD. Our objective was to estimate DSe and DSp of common tests to detect AGD and CGD in Scotland's farmed salmon under field conditions. Goals were to 1) to compare estimates for DSe and DSp of gross AGD scores, 2) provides estimates for DSe and DSp for histopathology for AGD and CGD, and PCR for AGD, and 3) provide the first estimates of DSe and DSp for gross gill scores for CGD.

2. Material and methods

2.1. Source populations and sampling strategy

Samples belong to a longitudinal prospective study investigating marine gill disease in cultured Atlantic salmon which took place between September 2018 and June 2020 on 8 production sites belonging to 6 different salmon producers in Scotland. On each site, two pens per site were sampled bi-weekly (weather and other circumstances allowing) from initial stocking into saltwater until these pens were harvested (6 sites) or COVID-19 related restrictions were imposed on the sites (2 sites). This led to 18–34 sampling points per site. Eight fish were visually assessed and sampled in each pen at each sampling point, of which 4 were lethally sampled to allow collection of gill tissue and 4 non-lethally sampled. Only the 4 lethally sampled fish per pen were considered in this study. Samples were taken by trained site staff or fish health representatives from each of the different companies involved in the project. All animal procedures were approved by the lead institute's Animal Ethics Committee.

2.2. Common tests for AGD and CGD

We evaluated the three most relevant tests for AGD and CGD currently in use by producers in Scotland; for AGD this included gross AGD scores, qPCR for *N. perurans* and histopathology. For CGD, gross PGD scores, gross total gill score, and histopathology were evaluated. Other putative pathogens that have been reported to be associated with CGD (Boerlage et al., 2020) were not considered in this work because their relationship to CGD is yet to be fully elucidated.

2.2.1. Gross gill scores

Three types of gross gill scores (i.e. gross AGD score, gross PGD scores and gross total gill scores) were recorded by site staff, who also performed the sample collection. All site staff who participated in the study were trained to score gills and collect samples according to the study protocol. Gill scoring competence and sample collection technique were evaluated once per quarter or more frequently where possible. It was not feasible to assess repeatability due to staff turnover within sites. All staff participating in scoring of gills were already familiar with gill assessment and scoring due to each company's policies of scoring gills on a weekly basis during the obligatory lice counts. Fish were selected at random from the pen and anaesthetised using Tricaine. Fish were lifted in random order one by one out of the anaesthetic bath. The operculum of the fish was opened and all hemibranchs were visually checked and scored according to Table 1. AGD scores are based on the worst gill arch, the protocol was based on the gill scoring guide by (Taylor et al., 2009b). PGD scores are also scored using the worst gill arch and are based on the syndrome proliferative gill disease, a non-specific term derived from examinations of gross gill lesions and histopathological lesions indicating there is proliferation in the gills (Mitchell et al., 2012b), that is

Table 1
Description of gross gill scores.

Score	AGD	PGD	Total gill
0	No sign of infection and healthy red colour	No sign of proliferative changes and healthy red colour	Gills appear healthy, with no visible lesions, abnormal colour, or excessive mucus
1	1 white spot, light scarring, or undefined necrotic streaking	Very slight thickening or very few lamellae affected	Focal lesion present on 1 or 2 gill arches only, 1–5% of total gill area affected
2	2–3 spots/small mucus patch	Frequent thickening but only affecting tips	More than 1 lesion. 5–25% of total gill area affected
3	Established thickened mucus patch or spot groupings up to 20% of gill area	Most lamellae have thickened tips, with some affected to more than 50% of lamellae length	25–50% of total gill area affected
4	Established lesions covering up to 50% of gill area	Most lamellae thickened progressing to more than 50% of lamellae length	50–75% of total gill area affected
5	Extensive lesions covering most of the gill surface (50%+)	Almost all lamellae affected along entire length	75–100% of total gill area affected
Focus	Worst gill arch	Worst gill arch	All gill surfaces

considered part of CGD (Boerlage et al., 2020; Herrero et al., 2018). Gross total gill scores make use of all gill surfaces and are a measure of how much of the total gill area is visually affected by anything abnormal.

2.2.2. Real-time RT-PCR for N. perurans

After gills were grossly scored, hemibranch 2–7 on the left side were swabbed and processed according to (Fringuelli et al., 2012). Briefly, gill swab samples in eNAT (COPAN diagnostics) were vortexed in a tube containing the transport media for 10 s and centrifuged quickly. Nucleic acids (DNA and RNA) were extracted from 500 µl of the transport media using Roche's MagNA Pure 96 instrument, MagNA Pure 96DNA and Viral NA large volume extraction kit, following the protocol Pathogen Universal 500. Extracted DNA and RNA were eluted in 50 µl final volume of the supplied kit elution buffer. 2.5 µl of extracted DNA and RNA were used in a duplex reaction together with an assay for elongation factor 1 alpha (EF1a, using Roche's Light Cycler Multiplex RNA Virus Master), in a total volume of 10 µl in each well. Primer and probe concentrations were 500 µM and 200 µM, respectively, and 130 µM and 100 µM for the EF1a assay. Samples were run in a single well for each sample and for each pathogen assay. Results were reported on a semi-quantitative scale with any results above 38 considered negative through to 1 (extremely positive) using cycle thresholds (Ct) observations. Individuals who processed the samples had no access to clinical information. Real-time RT-PCR is further referred to as "PCR."

2.2.3. Histopathology

After fish were swabbed, they were lethally anaesthetised using an overdose of Tricaine. The second gill arch on the right was removed and stored in 10% formal buffered saline. Samples were fixed in 10% neutral buffered formalin, routinely processed, and embedded in paraffin wax blocks, sectioned (3–5 µm), and stained with haematoxylin and eosin (H&E). Two specific criteria for AGD were recorded: "presence of amoeba" and "AGD related pathology." "Presence of amoeba" ranged from 0 (absence) to 1 (presence) and all cases of "presence of amoeba" were also positive for "AGD related pathology." "AGD related pathology" had more positive cases and was semi-quantitative (0 – 3), therefore, we used only "AGD related pathology" in our case definition for AGD, see Table 2.

For CGD, histopathology was based on a semi-quantitative system developed by (Mitchell et al., 2012b). The criteria included were "lamellar hyperplasia" (0−3), "lamellar fusion" (0−3), "cellular hypertrophy" (0−3), "cellular death" (0−3), "lamellar oedema" (0−3), "vascular changes" (0−3), and "inflammation" (0−3). These pathological changes in gills that are often associated with CGD (Noguera et al., 2019). The observed scores from these 7 criteria were aggregated to a single number that represented the histopathology score, see Table 2. The semi-quantitative ranges for AGD and CGD can be described as following: 0 = absence, 0.5 = minimal, 1 = mild, 2 = moderate, and 3 = severe. Individuals who processed the samples had no access to clinical information.

Table 2
Case definitions for AGD and CGD: cut-offs to determine a case.

Disease	Test	Less stringent	Medium stringent	Severe stringent
AGD	Gross AGD score	≥ 1	≥ 2	≥ 3
AGD	PCR: Ct value	< 38	< 23	< 20
AGD	Histopathology: presence of AGD pathology	≥ 1	≥ 2	≥ 3
CGD	Gross total gill score	≥ 1	≥ 2	≥ 3
CGD	Gross PGD score	≥ 1	≥ 2	≥ 3
CGD	Histopathology: histopathology score	≥ 4	≥ 7	≥ 11

2.3. Case definitions for AGD and CGD

Cases and non-cases of disease were distinguished using different cut-offs for tests according to Table 2. The cut-offs for gross gill scores were based on interviews with producers and represent perceived level of gill disease. For PCR, the cut-off for the less stringent case definition was based on the laboratory results designating a Ct value < 38 as positive. The other cut-offs for PCR were based on the sensitivity analysis (see below), i.e. running the Bayesian latent class models with different cut-offs and selecting cut-offs that were stable, with good DSe and DSp and of which the numbers of positive cases were similar to the number of cases for gross AGD score and histopathology results for the same level of stringency. The cut-offs for histopathology were for AGD the most sensible option considering the scale, and for CGD based on discussions with the pathologists that assessed the samples. We also tested a variety of cut-offs in the sensitivity analysis. Clinical significance of the three case definitions was based on histopathology (Mitchell et al., 2012a) and expert opinions, identifying the less stringent case definition to include cases of mild to high clinical significance, the medium stringent case definition cases of moderate to high clinical significance, and the severe stringent case definition only cases of high clinical significance.

2.4. Rationale for using latent class models

When there are doubts about whether the reference standard is perfect, estimates of DSe and DSp for the other tests will be flawed since they would be based on an imperfect standard (OIE, 2019). In our case, there are concerns whether histopathology can serve as a perfect reference standard. First, most AGD lesions are present in the dorsal or ventral gill extremities (Hytterød et al., 2018), but often we observed a sampler bias to collect the middle of the gill arch only. Second, gross AGD and PGD scores are based on the worst gill arch, and gross AGD scores have a better correlation with histopathology if this worst gill arch is used for histopathology compared to another gill arch (Adams et al., 2004). Between two of the tests, PCR and histopathology, there is a direct conflict in optimal sampling methods. Use of a gill swab to collect gill mucus for PCR analysis, makes the swabbed gill surface unsuitable to be used for histopathology, and thus these samples must be taken from different sections of gill tissue. A solution is to designate a specific gill arch to each test, which is common and has been done in this

study, but this may jeopardise the perfectness of histopathology as reference standard. Third, for CGD, one of the gross gill scores represented all gill surfaces. It was not possible to collect and score all, nor would this be a practice commonly done in the field.

2.5. Data management and analysis

Data management was done using R version 4.0.2 (R Core Team, 2018), RStudio version 1.3.1073 (RStudio Team, 2020), and the Tidyverse package (Wickham et al., 2019). Estimation of DSe and DSp of common tests for AGD and CGD were estimated through Bayesian latent class modelling using a "three tests in two or more populations" – approach (Branscum et al., 2005). Each site was considered a population, so there were 8 populations. Flat priors (beta 1,1) were used because no prior information was available on estimates of DSe and DSp for CGD, and of prevalence of both AGD and CGD on the 8 farms. Markov-chain Monte Carlo (MCMC) estimation was carried out using the R-package runjags (Denwood, 2016). For AGD with 3 chains, a burn-in period of 1000 iterations, a sample of 20,000 iterations, and a thin of 20 draws was used. For CGD we used 3 chains, a burn-in period of 1000 iterations, a sample of 40,000 iterations, and a thin of 20 draws. Initial values for both AGD and CGD for DSe were 0.5, 0.99 and 0.5 and for DSp 0.99, 0.75 and 0.99. All settings were kept the same for the three different case definitions and other tests run as part of the sensitivity analysis, unless mentioned. We have added a script for AGD severe stringent case definition in the supplements. We report the median, lower 95% and upper 95% boundaries. MCMC convergence was monitored using visual assessment of the trace plots, verifying that the Gelman-Rubin statistic was less than 1.05 (Gelman et al., 2013) and the effective sample size was greater than 1000. Missing data were appropriately identified (see example script S1 in the Supplements).

2.6. Sensitivity analysis

Robustness of outcomes were investigated by running different models varying different parts of the basic analysis and comparing the model outcome.

2.6.1. Model assumptions (including Period 1 and Period 2)

First, we investigated the model assumptions. The three main assumptions, also known as the Hui-Walter paradigm (Hui and Walter, 1980; Toft et al., 2005) were: 1) the tests are independent and only conditional on the infection status of animals, 2) the populations (farm A – H) have different prevalence of disease, and 3) DSe and DSp remain unchanged across populations (farm A – H). To test assumption 1, we ran the tests with and without conditional dependence between all combinations of tests, (i.e. conditional dependency between a) test 1 and test 2, b) test 1 and test 3, c) test 2 and test 3 and d) test 1, test 2 and test 3) as conditional dependence may affect the estimates (Gardner et al., 2000). We also split the outbreak into a naïve/ early infection period (hereafter referred to as Period 1) and a late/chronic infection period (hereafter referred to as Period 2) because the pathogenesis of disease, population dynamics and host susceptibility may change over time or because of disease. We did this by creating a two-month gap by removing two months of observation during the peak of infection. We compared the estimates for DSe and DSp in Period 1 and Period 2. We also changed these periods to have longer or shorter "tails" (i.e. removing early and late observations when disease was close to absent). To confirm that prevalence on each farm was different under assumption 2, we estimated prevalences per farm. To test if the data met assumption 3, we removed one farm at a time, to make sure there was no bias effect from a particular farm.

2.6.2. Cut-offs

In addition to checking the assumptions, we varied for AGD the cut-off between disease and non-disease of the Ct. values for the PCR analysis. We did not examine the cut-off for the gill scores, as the scores vary between 0 and 5 and the designation of the different levels of stringency was the only possibility. For AGD we also removed observations that were produced using different extraction methods for PCR.

2.6.3. Priors

For both CGD and AGD we investigated the influence of priors by assessing the effect of using beta(2,2), beta(2,8), and beta(8,2) on the posterior density instead of the flat prior beta(1,1) used in the main model.

2.7. Reporting

We have followed the recommendations for reporting standards for test accuracy studies STRADAS-Aquatic (Gardner et al., 2016) and STARD-BLCM (Kostoulas et al., 2017).

3. Results

3.1. Descriptive

Between stocking and harvesting of the fish groups on each site, there was a peak of AGD severity in late summer and autumn after which severity declined (Fig. 1). Most farms had one peak between stocking and harvesting, but farm D and H had or were heading to a second outbreak, resulting to a second peak in late summer and autumn. It was interesting that the different tests for AGD within farm had peaks at different times, even though the peaks were close in time. Test results for AGD and CGD followed a similar pattern in the early phase before disease peaked, except for farm F. After the disease peaked, test results for CGD kept on increasing or stayed at a stagnant level for 3 of the sites.

The number of observations per test combination per farm for the three different case definitions can be found in Table 3. Total number of observations were the same for the three case definitions within farm because the only difference was that cases became non-cases as stringency of the case definition increased, and thus there were least cases for the severe stringent case definitions. For farm A, C, D, and G there was a difference in total number of observations between tests for AGD and tests for CGD because of missing test results. There were 1656 observations for all tests for AGD, 31 missing observations for histopathology (0, 3, 21, 1, 1, 2, 0, 3 for farm A-H resp.) and 43 missing observations for gross AGD scores (0, 9, 0, 8, 16, 4, 5, 1). For CGD there were 1650 observations for all tests, 32 missing observations for gross total gill score (8, 0, 0, 16, 0, 0, 8), 42 missing observations for histopathology (0, 3, 29, 2, 1, 2, 2, 3) and 43 missing observations for both gross PGD score and gross total histopathology score (0, 9, 0, 8, 16, 4, 5, 1; which are the same as missing gross AGD scores). Reasons behind these missing observations included staff forgetting to take a sample, scoring form being misplaced, wind gusts blew samples into the loch or sample quality was insufficient for processing. We did not suspect or detect any bias resulting from missing samples. For the severe stringent case definition, the PCR test had the most positive observations from the AGD tests, and gross total gill score from the CGD tests, but the differences between the number of positive observations among the three tests were small for AGD.

3.2. Estimates for DSe and DSp

The posterior distribution estimates can be found in Table 4. For AGD, gross AGD scores had a median DSe of 0.53 (0.48–0.57 lower to upper 95% credible interval), which was similar for a medium stringent case definition, but only 0.14 (0.07 – 0.22) for a severe stringent case definition. For a medium stringent case definition, DSe for PCR scored the highest with 0.92 (0.86 – 0.99) and histopathology lowest with 0.23 (0.16 – 0.30) for a medium stringent case definition. DSp scored close to 1 for all tests and all stringencies, except for PCR at a medium stringent

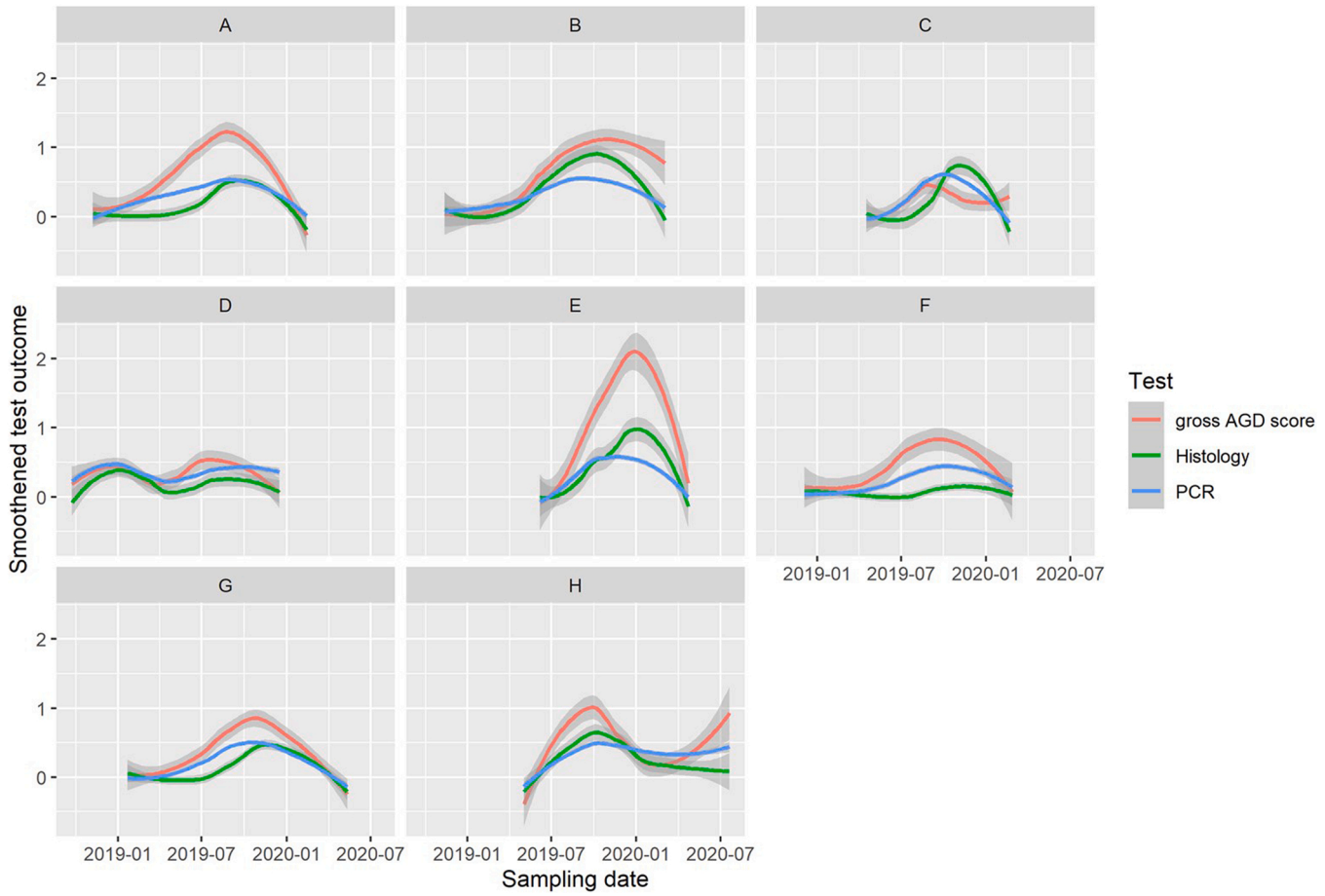


Fig. 1. Smoothened time series of outcomes of common tests for AGD for the 8 study farms. PCR test were based on Ct. values, which were inversed and normalized for visualization purposes.

case definition, which was 0.68 (0.64 – 0.73).

For CGD, DSe of histopathology was the lowest, at similar scores as for AGD. DSe of gross total gill score was highest at 0.9 (0.87 – 0.92) and 0.8 (0.64 – 0.99) for less and severe stringent case definitions respectively. DSe of gross PGD scores was in between, with 0.53 (0.46 – 0.64) for a medium stringent case definition.

3.3. Sensitivity analysis

Regarding assumption 1 on test independence, conditional dependency between tests was negligible, with the maximum median estimate being 0.08 for DSp with covariance between all 3 tests when using a less stringent case definition for Period 2, and the median of the median covariance estimates were < 0.001 for both AGD and CGD. There was a negligible effect on the estimates for DSe and DSp, see Supplements S2. The comparison of estimates of DSe and DSp for AGD and CGD between period 1, 2 and the entire dataset can be found in Fig. 3 and Table S8. Only small differences could be observed, except for the estimate for DSe for gross AGD score for AGD for a less stringent case definition of 0.59 (0.51 – 0.67) for Period 1 and 0.39 (0.31 – 0.46) for Period 2, and the DSp for gross total gill score for CGD for a less stringent case definition of 0.72 (0.66 – 0.78) for Period 1 and 0.42 (0.32 – 0.52) for Period 2. Reducing the size of Period 1 and 2 further for common tests for AGD resulted in small differences except for with the largest cut-off, which we attributed to reducing the dataset too much (see Supplement tests for AGD changing cut-offs to 0.3 in S3). We did not analyse this for CGD because the time series graph did not show one single peak for each farm (see Fig. 2).

Regarding assumption 2 on unequal prevalence, for AGD, estimated prevalence ranged from 0.34 (0.26 – 0.42) to 0.84 (0.76 – 0.91) for the less stringent case definition, 0.09 (0.05 – 0.14) to 0.42 (0.32 – 0.53) for the medium stringent case definition, and 0.02 (<0.01 – 0.08) to 0.41 (0.26 – 0.57) for the severe stringent case definition. For CGD, prevalence ranged from 0.18 (0.09 – 0.28) to 0.98 (0.09 – 1) for the less stringent case definition, 0.10 (0.06 – 0.14) to 0.48 (0.04 – 0.59) for the medium stringent case definition, and < 0.01 (<0.01 – 0.02) to 0.26 (0.02 – 0.36) for the severe stringent case definition.

Regarding assumption 3 of unchanged DSe and DSp across populations, there was little variation when leaving out one site for all estimates, compared to the estimates with all sites included. Among the few differences were for AGD in Period 2 for the less stringent case definition. DSe for histopathology increased to 0.67 (0.41 – 0.94) and 0.67 (0.54 – 0.81) when respectively site G or H were excluded, and for gross gill scores to 0.70 (0.49 – 1.00) and 0.73 (0.62 – 0.83) when respectively site G or H were excluded, DSp for PCR decreased to 0.35 (0.25 – 0.47) and 0.61 (0.54 – 0.68) when respectively site G or H were excluded. For CGD the only difference was also in Period 2 for the less stringent case definition for absence of site H, the estimate for DSp of gross total gill score increased to 0.78 (0.63 – 0.94). See supplement S4.

The cut-off for the PCR tests was based on the Ct. values. The less stringent cut-off was set at 38 Ct.'s which is based on analytic sensitivity and specificity testing by the commercial laboratory (beyond the scope of this research). Estimates for DSe and DSp for a range of cut-offs did not vary much for the medium stringent case definition, and a cut-off of 23 was used because for a higher cut-off the DSp estimate was lower. For the severe stringent case definition, we selected the highest Ct. value because it provided better estimates for DSe of the PCR test. There was no effect of varying cut-offs for the PCR test on estimates of DSe or DSp

Table 3
Cross-tabulated number of observations per test combination used in the analysis for AGD and CGD for the different case definitions.

Case definition				Less stringent								Medium stringent								Severe stringent							
Farm				A	B	C	D	E	F	G	H	A	B	C	D	E	F	G	H	A	B	C	D	E	F	G	H
Tests[a] for AGD	1	2	3																								
Number of test positive or negative for AGD	+	+	+	36	51	7	18	33	9	27	33	6	6	0	0	8	0	4	8	0	0	0	0	0	0	1	0
	+	+	–	7	27	16	25	8	4	8	25	2	17	6	5	1	1	1	4	0	2	1	0	2	0	0	1
	+	–	+	1	0	0	1	0	0	0	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
	+	–	–	1	3	0	0	0	3	1	0	0	2	0	0	0	0	1	2	0	0	0	0	0	0	0	0
	–	+	+	67	34	13	44	30	43	36	35	21	16	2	6	29	15	12	20	6	4	0	0	14	0	1	7
	–	+	–	72	51	22	95	7	38	42	90	98	78	44	109	30	37	68	73	72	78	45	68	42	34	53	65
	–	–	+	16	19	4	7	7	8	6	4	2	8	1	1	1	1	3	2	0	3	0	1	0	0	0	3
	–	–	–	71	56	73	25	42	77	155	58	142	114	82	94	58	128	186	137	193	154	89	146	69	148	220	171
Sum of all observations				271	241	135	215	127	182	275	247	271	241	135	215	127	182	275	247	271	241	135	215	127	182	275	247
Tests[a] for CGD	4	5	6																								
Number of test positive or negative for CGD	+	+	+	47	26	37	22	31	31	21	20	1	4	6	2	4	1	1	2	0	0	1	0	0	0	0	0
	+	+	–	1	6	1	10	1	9	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
	+	–	+	25	11	6	5	0	7	11	26	2	2	0	0	4	0	1	3	0	0	0	0	1	0	0	0
	+	–	–	3	17	9	11	2	6	6	3	4	4	5	6	4	1	1	6	0	0	0	2	1	0	0	1
	–	+	+	74	27	35	77	53	68	59	38	16	12	20	22	19	32	13	9	7	6	12	0	6	5	3	2
	–	+	–	3	2	1	7	3	23	3	4	3	4	10	7	2	12	3	1	4	2	3	0	1	3	2	1
	–	–	+	61	44	5	20	9	19	42	113	25	9	1	15	27	21	8	44	5	6	0	0	22	2	5	11
	–	–	–	49	108	33	46	28	19	122	42	212	206	85	146	67	114	237	182	247	227	111	196	96	172	255	232
Sum of all observations				263	241	127	198	127	182	265	247	263	241	127	198	127	182	265	247	263	241	127	198	127	182	265	247

[a] Test 1 = Gross AGD score, Test 2 = qPCR for *N. perurans*, Test 3 = Histopathology for AGD, Test 4 = Histopathology for CGD, Test 5 = Gross PGD Scores, Test 6 = Gross total gill score

for histopathology or gross AGD score (see Supplement S5). Two different extraction methods were used during the study due to COVID-19 related reagent shortages. Deleting the extraction method with the least observations, total nucleic acid (TNA) extraction in 96 samples, did not affect the estimates for histopathology or gross AGD score, except for an increase in DSe in Period 2 for a less stringent case definition for histopathology to 0.76 (0.60 – 0.90) and 0.63 (0.49 – 0.76) for a dataset without the samples and a dataset without the samples below Ct. = 27 for the different extraction method, and for gross AGD score to 0.69 (0.54 – 0.89) and 0.65 (0.52 – 0.77) respectively. The extraction method affected the estimate for DSe in Period 1 and 2 by reducing it to zero for a severe stringent case definition, which was the result of severe cases being dropped from the dataset, which could have been due to chance because the numbers were low. See supplement S6.

Lastly, the effect of changing priors had little effect on the estimates. There was for AGD a change when using prior beta(2,8) which was considered unrealistic because previous work by (Adams et al., 2004) indicated the DSe might be more towards a beta(8,2). For CGD, there was little variation as well, except for the DSp of PCR for a less stringent case definition in Period 2 when using the beta(2,8) prior, which was 0.50 (0.43 – 0.58). See supplement S7.

4. Discussion

In this study, we have estimated DSe and DSp of common tests to detect AGD in Scotland's farmed salmon under field conditions, and it is to our knowledge the first attempt to estimate DSe and DSp for gross gill scores for CGD. For both AGD and CGD, gross gill scores were excellent in designating all truly non-diseased fish (DSp close to 1 for all gross gill scores) and medium to good in designating truly diseased fish (DSe 0.5 – 0.9), implying there were additional false-negatives, but almost all positives were true positives. Out of the included tests, the PCR test had the best performance for AGD (high DSe and high DSp), for CGD the gross total gill score was most optimal. DSe of histology was low for both AGD and PGD, which we speculate was because the sample for histopathology was always taken from the second gill arch on the right rather than selecting the worst affected arch.

The 95% credible intervals of the estimates deviated from the median up to 0.3, indicating the accuracy is good. This was also the case for the medium and severe stringent case definitions, for which there were fewer observations in some of the categories (e.g. the category positive gross AGD score, negative qPCR for AGD and positive histopathology result for AGD) for the medium and severe stringent case definitions. A low number of observations in a category does not preclude a reliable estimate and this has been demonstrated in other studies (see e.g. Bates et al., 2020). The models for the medium and severe stringent case definitions converged, implying that the estimates are not unstable. Also, uncertainty resulting from a low number of cases in a category would be expressed in large credible intervals. In our study, these intervals were up to 0.3 from the median, therefore we considered the estimates to be robust, also for the clinically relevant case definitions, medium and severe stringent.

For gross AGD scores, estimates for DSp were similar between our study (0.82 (0.76–0.87)) and previous studies (0.75–0.91), but estimates of DSe were lower in our study (0.53 (0.48–0.57)) versus those previous studies (0.82–0.78) (Adams et al., 2004; Rozas et al., 2011). There can be many reasons for this difference. First, different methods were used to estimate the DSe and DSp which may have influenced the results. Second, case definitions were different. Our gross AGD score was based on the worst gill arch, whereas Adams et al. (2004) and Rozas et al. (2011) based them on all gill surface. Adams et al. (2004) looked at presence of gross gill lesions, but do not clarify if these were AGD-specific lesions or any gill lesions. Rozas et al. (2011) specify gross AGD pathology as pale gill and raised multifocal white mucoid patches which may be comparable to a score of 2 on the scale of our gross AGD score. Variability between protocols for gross AGD scores is common throughout the

Table 4
Posterior distributions estimates for common tests for AGD and CGD, by the different case definitions.

	Less			Medium			Severe		
	Median	Lower95	Upper95	Median	Lower95	Upper95	Median	Lower95	Upper95
AGD									
Dse[gross AGD score]	0.53	0.48	0.57	0.50	0.37	0.63	0.14	0.07	0.22
Dse[PCR]	0.96	0.94	0.99	0.92	0.86	0.99	0.83	0.68	0.97
Dse[histopathology AGD]	0.34	0.31	0.39	0.23	0.16	0.30	0.03	0.01	0.06
DSp[gross AGD score]	0.91	0.88	0.94	0.99	0.98	1.00	1.00	0.99	1.00
DSp[PCR]	0.87	0.78	0.96	0.68	0.64	0.73	0.81	0.76	0.86
DSp[histopathology AGD]	0.99	0.98	1.00	1.00	0.99	1.00	1.00	1.00	1.00
CGD									
DSe[gross PGD score]	0.82	0.76	0.87	0.53	0.46	0.64	0.45	0.34	0.56
DSe[gross total gill score]	0.90	0.87	0.92	0.81	0.73	0.91	0.80	0.64	0.99
DSe[histopathology CGD]	0.37	0.33	0.40	0.10	0.07	0.14	0.03	0.00	0.06
DSp[gross PGD score]	1.00	0.98	1.00	0.99	0.98	1.00	1.00	0.99	1.00
DSp[gross total gill score]	0.66	0.61	0.71	0.99	0.95	1.00	1.00	0.99	1.00
DSp[histopathology CGD]	0.89	0.86	0.91	0.98	0.97	0.99	1.00	0.99	1.00

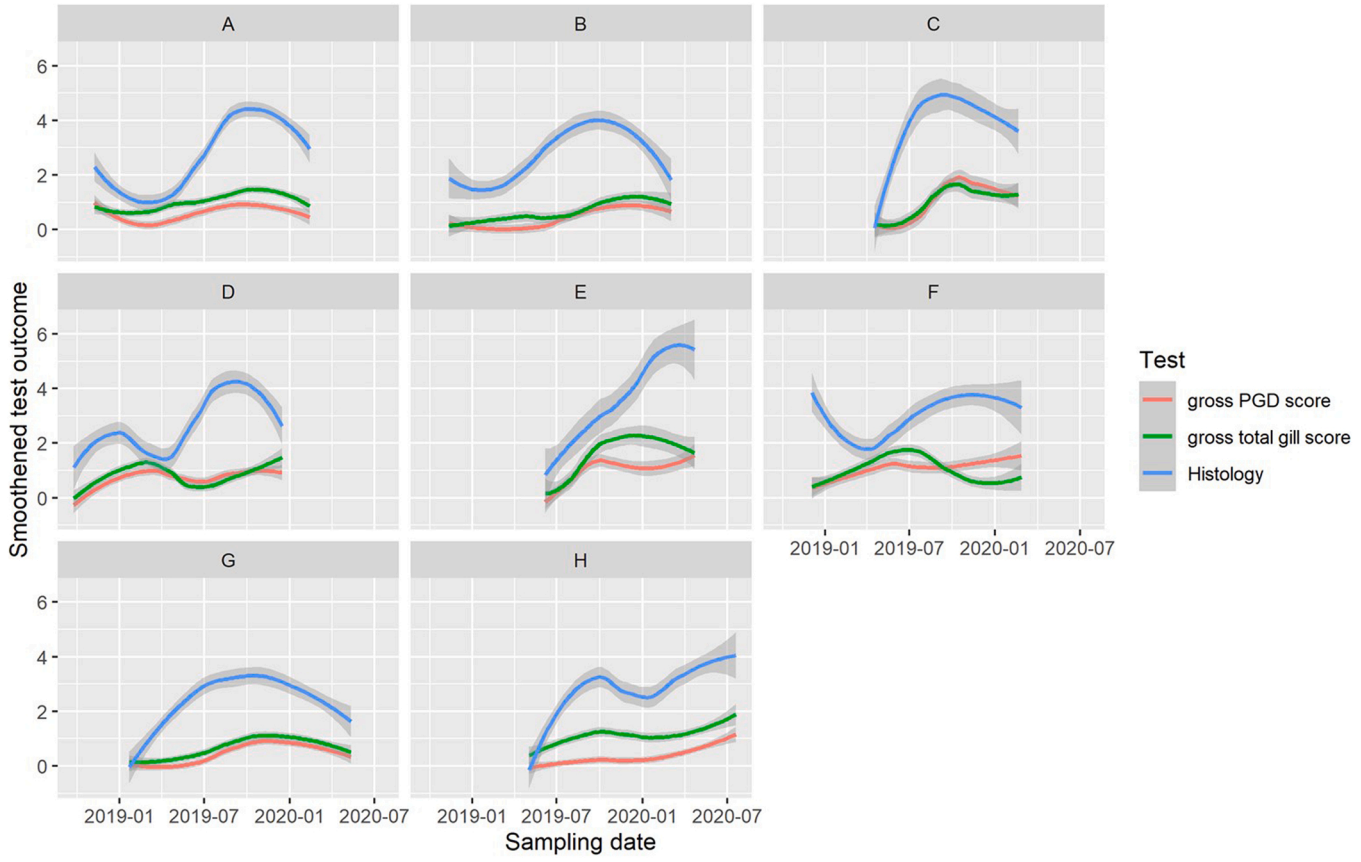


Fig. 2. Smoothened time series of outcomes of common tests for CGD for the 8 study farms.

salmon farming industry and compromises comparisons between scores captured by different studies or companies. In a study that investigated gross scores at different phases of AGD development, Hytterød et al. (2018) recommend that the type of gross AGD scores used should depend on the goal of sampling, i.e. for use of surveillance in the field based on the gill surface with the highest score, for monitoring treatment efficacy based on the number of free gill surfaces, and for laboratory studies when small differences need to be captured, based on more elaborate gill scores that take all gill surfaces into account. The scoring system used in our study was based on the worst gill arch, and therefore suitable for surveillance in the field. Other reasons for a lower DSe compared to previous studies could be the type of analysis used, as we did not use histopathology as reference standard, but instead as one of the tests. Furthermore, our study was done at a different time and in a different part of the world compared to the other studies, so that hosts and pathogens could be genetically different (English et al., 2019; Robledo et al., 2020) which may affect gross gill scores (Taylor et al., 2009a), and gross gill changes (such as those caused by environmental insults) are likely to vary between populations of farmed fish in different parts of the world. Our study also included different companies and different samplers, which may have affected the estimates, because sampler bias could affect test estimates despite all samplers being trained to follow the same protocols. We perceive including this bias in the estimate as an advantage because they reflect field estimates for the Scottish salmon industry better.

For PCR, our estimates of both DSe and DSp were similar to the ones

previously estimated by Rozas et al. (2011). Their estimates were within our credible intervals for the less stringent case definition, which is most comparable to the presence/absence cut-off used by Rozas et al. (2011), even though they used a PCR and we a qPCR test. The meaning of varying the cut-off Ct. scores for our medium and severe stringent case definition may not be relevant, because Ct. scores are semi-quantitative values that cannot be directly compared between assays or different tests. Furthermore, differences between Ct. values are not linear, and therefore interpreting the meaning of different cut-off values is difficult, except when compared within assay.

Histopathology is an important tool commonly used to diagnose marine gill diseases. Histopathology has in the past been used as reference standard, therefore no estimates for DSe or DSp for histopathology were available. Our study found that estimates for DSp for histopathology were close to one, i.e. there were almost no false positives, but estimates for DSe were equally poor for AGD as CGD, ranging from 0.03 for the severe stringent case definitions to 0.35 and 0.37 for the less stringent case definitions, implying there were many false negatives. We speculate that the main reason for these false negatives may have been our sampling protocol of sampling the second gill arch on the right for histopathology. The worst gill arch is recommended for field sampling (Bustos et al., 2011; Downes, 2017; Slinger et al., 2020), but using the second gill arch on a previously defined side of the fish (left or right) is common practice (Adams et al., 2004; Bustos et al., 2011; Clark and Nowak, 1999; Hellebø et al., 2017; Mo et al., 2015), and a useful tool to reduce sampling bias when the worst gill arch cannot be selected because, for example, of a conflict over several tests competing for the same gill material, or when in doubt the sampler will select the worst gill arch accurately. Reason that the second gill arch became popular as a set sampling arch may be that AGD is often observed first on the second gill arch (Hytterød et al., 2018). However, when focal gill insults or abnormalities are located elsewhere on the gills, false negatives may arise. Comparisons with test results that are based on the worst gill arch, such as the gross gill scores for AGD and PGD, may lead to discrepancies and poor correlation. At clinically relevant gross AGD score of three, almost all gill surfaces have some macroscopic changes compatible with AGD (Hytterød et al., 2018), therefore it could be argued that the less or medium stringent case definition we used for histopathology might be more comparable to the severe stringent case definition used for AGD scores. Although this would lead to capturing more true positives, it would also lead to including more false positives, and reducing the DSp. To our knowledge it has not been defined how focal CGD is established on the gills in clinically relevant cases, but when considering the multi-factorial and non-specific nature of the disease, there might be much variation between cases. Interestingly, Adams et al. (2004) found only a small increase in DSe of gross AGD scores of 0.78–0.80 when using histopathology as reference standard based on the worst gill arch instead of the second gill arch on one site. Reason may be a difference in AGD presentation between Scotland and Tasmania. Our study could be followed up by estimating DSe of histopathology for both AGD and CGD between the worst and the second gill arch, and to obtain more information on the distribution of gill lesions throughout the gills for CGD.

To our knowledge, test characteristics for gross gill scores for CGD have not been estimated before. It was not surprising that the gross total gill score performed best, because it is an unspecific and inclusive score which reflects best our current understanding and definition of CGD. Gross PGD scores have been found ineffective in assessing gill health status in another study in Scotland (Król et al., 2020), which is not entirely similar our findings. Our estimate for DSp was about 1, indicating that all positive fish were true positives, and the DSe of 0.45 indicates that some negatives were false negatives (the exact proportion of false negatives depends on the prevalence (Dohoo et al., 2009)).

Different levels of variation are absorbed in the estimates. First, there was much variation in field circumstances. Samples were taken at different times at different sites by different samplers, implying that different levels of disease severity and differences between samplers, sites, location, and companies are embedded in the estimate and credible intervals. If such a study would be based on a challenge trial in a controlled environment with all samples taken at the same time, credible intervals may have been smaller. However, the estimates for DSe and DSp would only be applicable to the disease severity at the moment of sampling and under similar controlled conditions, whilst the estimates resulting from this study can be generalized for most field conditions in Scotland. Secondly, whilst AGD is caused by a single pathogen and very clearly defined, CGD is not. CGD is likely multifactorial, and no single putative pathogen or single type of environmental insult is consistently associated with cases, implying that different cases in time and space were likely different in aetiology and therefore gross and histopathological expression, on top of existing variation explained earlier in this paragraph. There are two reasons why it is important to emphasize this. First from an analytical standpoint, such variation can induce correlation between tests which can bias results. To mitigate that, dependencies between tests needs to be investigated (Pepe and Janes, 2007), which showed no dependency between tests in our study. Secondly, estimates are only as good as the case definitions are. The estimates from this study are a starting point and may become less useful when we understand and define CGD better and develop tests with a higher sensitivity and specificity. For now, these estimates are providing a base understanding of the meaningfulness of test results that can be used in the field in the absence of better estimates.

Estimates for Period 1 and 2 show that there were small differences in estimates of DSe and DSp in the beginning (before the peak) of the outbreak, compared to the end (after the peak) of the outbreak. However, these were almost always not significant according to the credible intervals (Fig. 3, Table S8 in the Supplements), and because in field situations it may not always be clear if individual fish are in a chronic or initial phase, therefore the overall estimates in Table 4 should be used for general purposes. The estimates of Period 1 and 2 in Table S8 provide additional understanding whether the test might perform slightly better or worse in early stages or under chronic stages of gill disease. These differences in estimates of before and after the peak may have arisen because of temporal changes in gill pathology due to the seasonality in disease progression and putative causal insults (Adams and Nowak, 2004; Gunnarsson et al., 2017). Furthermore, Figs. 1 and 2 show that CGD cases often remain more severe or chronic after the peak of infection of AGD, and thereby potentially causing more noise to histopathology and gill scoring of AGD. These figures also suggest that AGD was most likely part of the CGD pathology because the disease initiation of AGD and CGD commence synchronously. The exception was pen F, for which the peak in both gross gill scores was earlier than the peak of tests for AGD, indicating that AGD might not have been the part of the initial gross gill pathology.

As for any disease in any husbandry system, developing a surveillance strategy is important, and understanding of performance of tests included is essential. The tests compared in this study are common tests used in surveillance for marine gill diseases. All tests have a different purpose and hence should be used complementary. Gross gill scores can provide producers with a presumptive diagnosis and are an essential starting point (Hellebø et al., 2017). They are rapid and affordable and can be used in an active surveillance strategy when combined with the obligatory sequential lice counts. This way they can provide regular insight into the gill status of stock without additional disturbance of fish. In this role, gross gill scores have become important in health management of most salmon farms in Scotland. This study showed that almost all positive gill scores for AGD and CGD are most likely true positives when scoring is performed by properly trained individuals, but some negatives are false negatives. Knowing this information helps understanding what the gill scores mean, and when multiple fish are sampled can still provide the producer with a sound insight of the disease status of their stock. PCR tests can be used to confirm presence of specific pathogens. For CGD, PCR test may not be meaningful for a complete diagnosis, because most putative pathogens may be part of a

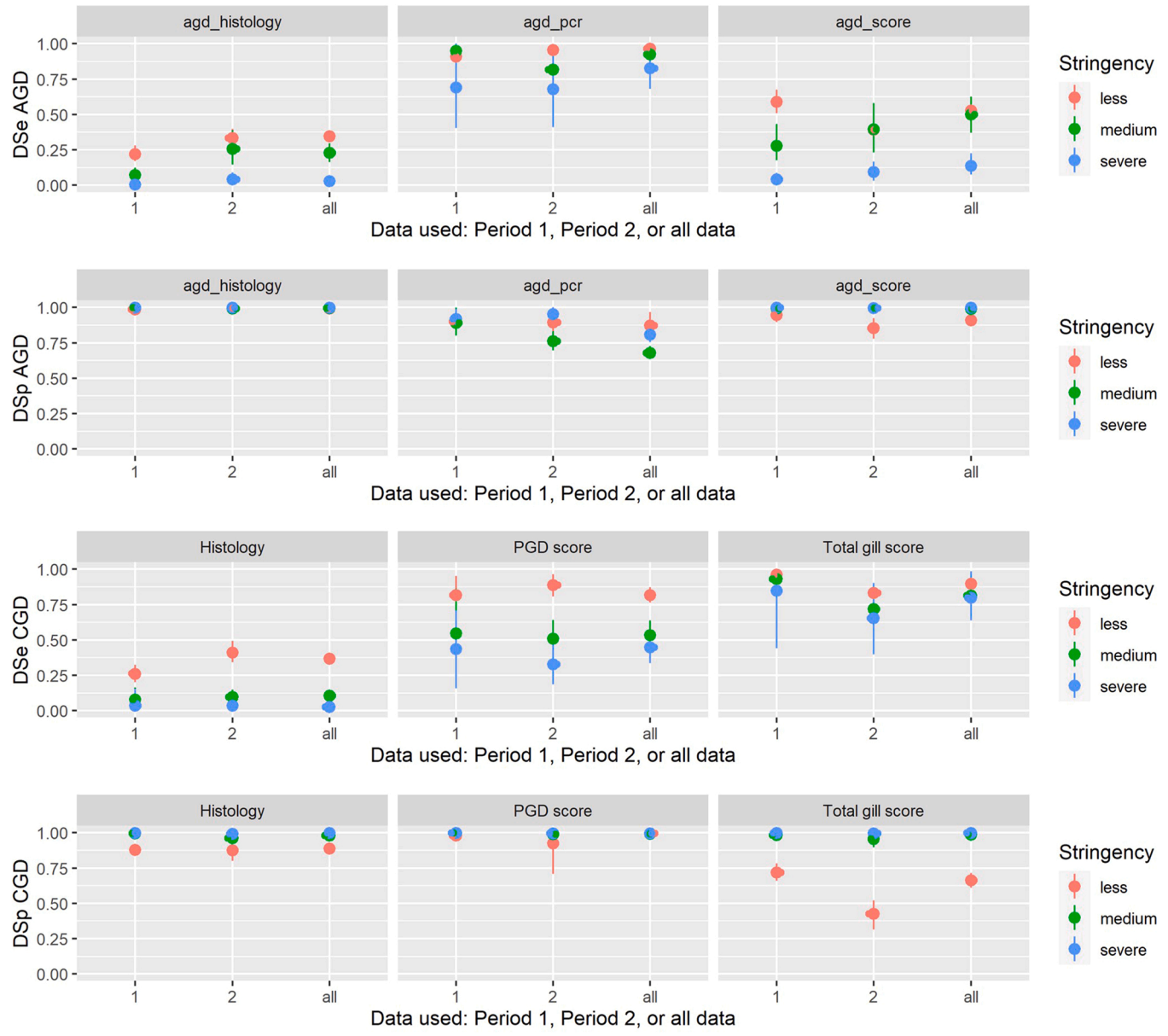


Fig. 3. Estimates for DSe and DSp for common tests for AGD and CGD for Period 1 and Period 2, for a less, medium, and severe stringent case definition.

CGD diagnosis, but not in all cases. For AGD, presence of *N. perurans* can be detected by PCR and used to confirm AGD. The preferred method of sampling for a PCR test is using a gill swab, which is non-lethal (Downes et al., 2017), and therefore preferred over lethal tests. Histopathology is a lethal test, which can be used as a second measure for confirmation. Histopathology provides much more information, such as which insults may be causing CGD, or it could give insight into disease progression or severity. Using the different tests while keeping in mind the test characteristics enables development of effective surveillance strategies which will enhance health management and eventually sustainability.

Acknowledgements

We are very appreciative for the commitment of all involved in the (field) work for this project, with special thanks to producer partners at Loch Duart Ltd, Cooke Aquaculture Scotland Ltd, MOWI Scotland, Grieg seafood Shetland, the Scottish Salmon Company, and Wester Ross Salmon. This work was supported by the Sustainable Aquaculture Innovation Centre, UK grant SL_2017_07.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.prevetmed.2022.105654.

References

Adams, M., Ellard, K., Nowak, B., 2004. Gross pathology and its relationship with histopathology of amoebic gill disease (AGD) in farmed Atlantic salmon, *Salmo salar* L. J. Fish Dis. 27, 151–161. https://doi.org/10.1111/j.1365-2761.2004.00526.x.

Adams, M.B., Nowak, B.F., 2001. Distribution and structure of lesions in the gills of Atlantic salmon, *Salmo salar* L., affected with amoebic gill disease. J. Fish Dis. 24, 535–542. https://doi.org/10.1046/j.1365-2761.2001.00330.x.

Adams, M.B., Nowak, B.F., 2004. Sequential pathology after initial freshwater bath treatment for amoebic gill disease in cultured Atlantic salmon, *Salmo salar* L. J. Fish Dis. 27, 163–173. https://doi.org/10.1111/j.1365-2761.2004.00531.x.

Bates, A., Laven, R., O'Brien, R., Liggett, S., Griffin, F., 2020. Estimation of the sensitivity and specificity of four serum ELISA and one fecal PCR for diagnosis of paratuberculosis in adult dairy cattle in New Zealand using Bayesian latent class analysis. Prev. Vet. Med. 185, 105199 https://doi.org/10.1016/j.prevetmed.2020.105199.

Bloecher, N., Powell, M., Hytterød, S., Gjessing, M., Wiik-Nielsen, J., Mohammad, S.N., Johansen, J., Hansen, H., Floerl, O., Gjevre, A.G., 2018. Effects of cnidarian

biofouling on salmon gill health and development of amoebic gill disease. PLoS One 13, 1–18. https://doi.org/10.1371/journal.pone.0199842.
Boerlage, A.S., Ashby, A., Herrero, A., Reeves, A., Gunn, G.J., Rodger, H.D., 2020. Epidemiology of marine gill diseases in Atlantic salmon (*Salmo salar*) aquaculture: a review. Rev. Aquac. https://doi.org/10.1111/raq.12426.
Branscum, A.J., Gardner, I.A., Johnson, W.O., 2005. Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. Prev. Vet. Med. 68, 145–163. https://doi.org/10.1016/j.prevetmed.2004.12.005.
Bustos, P.A., Young, N.D., Rozas, M.A., Bohle, H.M., Ildefonso, R.S., Morrison, R.N., Nowak, B.F., 2011. Amoebic gill disease (AGD) in Atlantic salmon (*Salmo salar*) farmed in Chile. Aquaculture 310, 281–288. https://doi.org/10.1016/j.aquaculture.2010.11.001.
Clark, A., Nowak, B.F., 1999. Field investigations of amoebic gill disease in Atlantic salmon, *Salmo salar* L., in Tasmania. J. Fish. Dis. 22, 433–443. https://doi.org/10.1046/j.1365-2761.1999.00175.x.
Denwood, M.J., 2016. runjags: An R package providing interface utilities, model templates, parallel computing methods and additional distributions for MCMC models in JAGS. J. Stat. Softw. 71, 1–25. https://doi.org/10.18637/jss.v071.i09.
Dohoo, I.R., Martin, W.E., Stryhn, H., 2009. Veterinary Epidemiology Research, second ed. VER Inc., Charlottetown.
Downes, J., Henshilwood, K., Collins, E., Ryan, A., O'Connor, I., Rodger, H., MacCarthy, E., Ruane, N., 2015. A longitudinal study of amoebic gill disease on a marine Atlantic salmon farm utilising a real-time PCR assay for the detection of *Neoparamoeba perurans*. Aquac. Environ. Interact. 7, 239–251. https://doi.org/10.3354/aei00150.
Downes, J.K., 2017. Gill Disease in Finfish Aquaculture with Particular Emphasis on Amoebic Gill Disease (PhD thesis). Galway/Mayo Institute of Technology, Galway.
Downes, J.K., Rigby, M.L., Taylor, R.S., Maynard, B.T., MacCarthy, E., O'Connor, I., Marcos-Lopez, M., Rodger, H.D., Collins, E., Ruane, N.M., Cook, M.T., 2017. Evaluation of non-destructive molecular diagnostics for the detection of *Neoparamoeba perurans*. Front. Mar. Sci. 4, 61. https://doi.org/10.3389/FMARS.2017.00061.
English, C.J., Tyml, T., Botwright, N.A., Barnes, A.C., Wynne, J.W., Lima, P.C., Cook, M. T., 2019. A diversity of amoebae colonise the gills of farmed Atlantic salmon (*Salmo salar*) with amoebic gill disease (AGD). Eur. J. Protistol. 67, 27–45. https://doi.org/10.1016/j.ejop.2018.10.003.
Fringuelli, E., Gordon, A.W., Rodger, H., Welsh, M.D., Graham, D.A., 2012. Detection of *Neoparamoeba perurans* by duplex quantitative taqman real-time PCR in formalin-fixed, paraffin-embedded Atlantic salmonid gill tissues. J. Fish Dis. 35, 711–724. https://doi.org/10.1111/j.1365-2761.2012.01395.x.
Gardner, I.A., Stryhn, H., Lind, P., Collins, M.T., 2000. Conditional dependence between tests affects the diagnosis and surveillance of animal diseases. Prev. Vet. Med. 45, 107–122. https://doi.org/10.1016/S0167-5877(00)00119-7.
Gardner, I.A., Whittington, R.J., Caraguel, C.G.B., Hick, P., Moody, N.J.G., Corbeil, S., Garver, K.A., Warg, J. v, Arzul, I., Purcell, M.K., Crane, M.S.J., Waltzek, T.B., Olesen, N.J., Lagno, A.G., 2016. Recommended reporting standards for test accuracy studies of infectious diseases of finfish, amphibians, molluscs, and crustaceans: The STRADAS-Aquatic checklist. Dis. Aquat. Org. 118, 91–111. https://doi.org/10.3354/dao02947.
Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., Rubin, D., 2013. Bayesian Data Analysis. CRC Press, Boca Raton.
Gunnarsson, G.S., Karlsbakk, E., Blindheim, S., Plarre, H., Imsland, A.K., Handeland, S., Sveier, H., Nylund, A., 2017. Temporal changes in infections with some pathogens associated with gill disease in farmed Atlantic salmon (*Salmo salar* L). Aquaculture 468, 126–134. https://doi.org/10.1016/j.aquaculture.2016.10.011.
Hellebø, A., Stene, A., Aspehaug, V., 2017. PCR survey for *Paramoeba perurans* in fauna, environmental samples and fish associated with marine farming sites for Atlantic salmon (*Salmo salar* L.). J. Fish Dis. 40, 661–670. https://doi.org/10.1111/jfd.12546.
Herrero, A., Thompson, K.D., Ashby, A., Rodger, H.D., Dagleish, M.P., 2018. Complex gill disease: an emerging syndrome in farmed Atlantic salmon (*Salmo salar* L.). J. Comp. Pathol. 163, 23–28. https://doi.org/10.1016/j.jcpa.2018.07.004.
Hjeltnes, B., Bornø, G., Jansen, M.D., Haukaas, A., Walde, C., 2017. The health situation in Norwegian aquaculture 2016.
Hui, S.L., Walter, S.D., 1980. Estimating the error rates of diagnostic tests. Biometrics 36 (1), 167–171. https://doi.org/10.2307/2530508.
Hytterød, S., Kristoffersen, A.B., Darrud, M., Kolstø, S., Mo, T.A., Blindheim, S.H., Andersen, L., 2018. Standardisering av AGD-gjellescore Enhetlig gjellescoring basert på data fra eksperimentelle forsøk og oppdrettsanlegg for laks. Vet. Rapp. 19, 23.
Kostoulas, P., Nielsen, S.S., Branscum, A.J., Johnson, W.O., Dendukuri, N., Dhand, N.K., Toft, N., Gardner, I.A., 2017. STARD-BLCM: standards for the reporting of diagnostic accuracy studies that use Bayesian latent class models. Prev. Vet. Med. 138, 37–47. https://doi.org/10.1016/j.prevetmed.2017.01.006.
Król, E., Noguera, P., Shaw, S., Costelloe, E., Gajardo, K., Valdenegro, V., Bickerdike, R., Douglas, A., Martin, S.A.M., 2020. Integration of transcriptome, gross morphology and histopathology in the gill of sea farmed Atlantic salmon (*Salmo salar*): lessons from multi-site sampling. Front. Genet. 11 https://doi.org/10.3389/fgene.2020.00610.
Mardones, J.I., Paredes, J., Godoy, M., Suarez, R., Norambuena, L., Vargas, V., Fuenzalida, G., Pinilla, E., Artal, O., Rojas, X., Dorantes-aranda, J.J., Lee, K.J., Anderson, D.M., Hallegraeff, G.M., 2021. Disentangling the environmental processes responsible for the world' s largest farmed fish-killing harmful algal bloom: Chile, 2016. Sci. Total Environ. 766, 144383 https://doi.org/10.1016/j.scitotenv.2020.144383.
Matthews, C.G.G., Richards, R.H., Shinn, A.P., Cox, D.I., 2013. Gill pathology in Scottish farmed Atlantic salmon, *Salmo salar* L., associated with the microsporidian *Desmozoon lepeophtherii* Freeman et Sommerville, 2009. J. Fish Dis. 36, 861–869. https://doi.org/10.1111/jfd.12084.
Mitchell, S.O., Rodger, H.D., 2011. A review of infectious gill disease in marine salmonid fish. J. Fish Dis. 34, 411–432. https://doi.org/10.1111/j.1365-2761.2011.01251.x.
Mitchell, S.O., Baxter, E.J., Holland, C., Rodger, H.D., 2012a. Development of a novel histopathological gill scoring protocol for assessment of gill health during a longitudinal study in marine-farmed Atlantic salmon (*Salmo salar*). Aquac. Int. 20, 813–825. https://doi.org/10.1007/s10499-012-9504-x.
Mitchell, S.O., Baxter, E.J., Holland, C., Rodger, H.D., 2012b. Development of a novel histopathological gill scoring protocol for assessment of gill health during a longitudinal study in marine-farmed Atlantic salmon (*Salmo salar*). Aquac. Int. 20, 813–825. https://doi.org/10.1007/s10499-012-9504-x.
Mo, T.A., Hytterød, S., Olsen, A.B., Hansen, H., 2015. Utvikling av amøbegjellesykdom (AGD) hos laks i tre oppdrettsanlegg i 2013-2014. Vet. Rapp. 9.
Noguera, P., Olsen, A.B., Hoare, J., Lie, K.I., Marcos-Lopez, M., Poppe, T.T., Rodger, H., 2019. Complex gill disorder (CGD): a histopathology workshop report. Bull. Eur. Assoc. Fish. Pathol. 39, 172–176.
OIE, 2019. Chapter 1.1.2. Principles and methods of validation of diagnostic assays for infectious diseases, Manual of Diagnostic Tests for Aquatic Animals. ⟨www.oie.int/fileadmin/Home/eng/Health_standards/aahm/current/chapitre_validation_diagnostics_assays.pdf⟩ (Accessed 9 July 2021).
Pepe, M.S., Janes, H., 2007. Insights into latent class analysis of diagnostic test performance. Biostatistics 8 (2), 474–484. https://doi.org/10.1093/biostatistics/kxl038.
R Core Team, 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org/ (Accessed 9 July 2021).
Robledo, D., Hamilton, A., Gutiérrez, A.P., Bron, J.E., Houston, R.D., 2020. Characterising the mechanisms underlying genetic resistance to amoebic gill disease in Atlantic salmon using RNA sequencing. BMC Genom. 21, 1–11. https://doi.org/10.1101/2020.12.03.410464.
Rodger, H., 2013. Amoebic gill disease (AGD) in farmed salmon (*Salmo salar*) in Europe. Fish Vet. J. 14, 16–26.
Rodger, H.D., 2007. Gill disorders: an emerging problem for farmed Atlantic salmon (*Salmo salar*) in the marine environment? Fish. Vet. J. 9, 38–48.
Rodger, H.D., Henry, L., Mitchell, S.O., 2011. Non-infectious gill disorders of marine salmonid fish. Rev. Fish. Biol. Fish. 21, 423–440. https://doi.org/10.1007/s11160-010-9182-6.
Rozas, M., Bohle, H., Ildefonso, R., Bustos, P., 2011. Development of PCR assay for detection of *Neoparamoeba perurans* and comparison of histological diagnosis. Bull. Eur. Assoc. Fish. Pathol. 31, 211–218.
RStudio Team, 2020. RStudio: integrated development for R. RStudio, PBC, Boston, MA. www.rstudio.com/ (Accessed 9 July 2021).
Slinger, J., Adams, M.B., Wynne, J.W., 2020. Comparison of bacterial diversity and distribution on the gills of Atlantic salmon (*Salmo salar* L.): an evaluation of sampling techniques. J. Appl. Microbiol. https://doi.org/10.1111/jam.14969.
Taylor, R.S., Kube, P.D., Muller, W.J., Elliott, N.G., 2009a. Genetic variation of gross gill pathology and survival of Atlantic salmon (*Salmo salar* L.) during natural amoebic gill disease challenge. Aquaculture 294, 172–179. https://doi.org/10.1016/j.aquaculture.2009.06.007.
Taylor, R.S., Muller, W.J., Cook, M.T., Kube, P.D., Elliott, N.G., 2009b. Gill observations in Atlantic salmon (*Salmo salar* L.) during repeated amoebic gill disease (AGD) field exposure and survival challenge. Aquaculture 290, 1–8. https://doi.org/10.1016/j.aquaculture.2009.01.030.
Toft, N., Jørgensen, E., Højsgaard, S., 2005. Diagnosing diagnostic tests: evaluating the assumptions underlying the estimation of sensitivity and specificity in the absence of a gold standard. Prev. Vet. Med. 68 (1), 19–33. https://doi.org/10.1016/j.prevetmed.2005.01.006.
Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the tidyverse. J. Open Source Softw. 4, 1686. https://doi.org/10.21105/joss.01686.