Title	The precision and robustness of published protocols for disc diffusion assays of antimicrobial agent susceptibility: an inter-laboratory study
Author(s)	Nic Gabhainn, Saoirse; Curtin, John; Smith, Pete
Publication Date	2004-02-02
Publication Information	Nic Gabhainn, S., Bergh, O., Dixon, B., Donachie, L., Carson, J., Coyne, R., et al. (2004). The precision and robustness of published protocols for disc diffusion assays of antimicrobial agent susceptibility: an inter-laboratory study. Aquaculture, 240(1-4), 1-18.
Publisher	Elsevier
Link to publisher's version	http://dx.doi.org/10.1016/j.aquaculture.2004.02.004
Item record	http://hdl.handle.net/10379/2316

ELSEVIER

The precision and robustness of published protocols for disc diffusion assays of antimicrobial agent susceptibility: an inter-laboratory study

Saoirse Nic Gabhainn[a], Øivind Bergh[b], Beverly Dixon[c],
Linda Donachie[d], Jeremy Carson[d], Rosie Coyne[b], John Curtin[e],
Inger Dalsgaard[f], Amedeo Manfrin[g], Grace Maxwell[h], Pete Smith[e,*]

[a] Department of Health Promotion, NUI, Galway, Galway, Ireland
[b] Division of Fish Health, Institute of Marine Research, Bergen, Norway
[c] Department of Biological Sciences, California State University, 25800 Carlos Bee Boulevard,
Hayward, CA 94542, USA
[d] Fish Health Unit, Department of Primary Industries, Water and Environment, P.O. Box 46,
Kingsmeadows, Launceston, Tasmania 7249, Australia
[e] Department of Microbiology, NUI, Galway, Galway, Ireland
[f] Danish Institute for Fisheries Research, Fish Disease Laboratory, Stigbøjlen 4,
DK-1870 Frederiksberg C, Denmark
[g] Fish Pathology Laboratory, Istituto Zooprofilattico delle Venezie, Via Romea 14/a, 35020 Legnaro, Padua, Italy
[h] Aquaculture Section, DFO, Gulf Fisheries Centre, 343 Ave. Universite, Moncton, NB E1C 9B6, Canada

Abstract

The precision of the disc diffusion protocols previously published by Alderman and Smith (Aquaculture 196 (2002) 211) was analysed in a seven-laboratory trial using *Escherichia coli* ATCC 25922 as the test strain. Discs containing seven antimicrobial agents were employed and 2899 zone size measurements were generated. The total data generated in the trial was used to quantify the intra- and inter-laboratory precisions. The study design also facilitated the investigation of the influence of the source of media and the source of discs on zone sizes. A smaller two-laboratory trial was employed to investigate the influence of incubation time of zone size.

The intra-laboratory precision was relatively high with the mean of the coefficients of variation calculated for each laboratory and each agent being 4.7%. In contrast, the inter-laboratory precision was very much lower with the mean of the values for each agent being 11.1%. Significant influences on zone size were detected for all three parameters of the protocol. Media source effects were

---

* Corresponding author.
  E-mail address: peter.smith@nuigalway.ie (P. Smith).

particularly notable with respect to oxytetracycline and oxolinic acid discs, disc source effects with respect to ampicillin and sulphamethoxazole/trimethoprim discs and incubation times with the ampicillin and amoxycillin discs. ANOVA analysis of the total data set confirmed that inter-laboratory variation was the major factor influencing the low precision of the protocol.

The overall precision of the protocols used here was found to be significantly lower than that implied by the control limits associated with the same bacterium in other validated disc diffusion protocols. The implications of these results, for the further development of the protocols under investigation, are discussed.

© 2004 Published by Elsevier B.V.

1. Introduction

The development of any method or set of methods, for determining the susceptibility of bacteria to antimicrobial agents is, necessarily, a long, sequential process. The initial steps in this process are the formulation of the experimental protocols, their validation and the establishment of quality control compliance criteria to be applied to their routine use. The final step, which can only be approached after the initial steps have been completed, is the development and validation of interpretive schemes that allow the establishment of the clinical meaning of the laboratory data generated by the protocols.

Alderman and Smith (2001) published a set of protocols for the laboratory determination of the susceptibility to antimicrobial agents of bacteria associated with fish diseases that had been developed by a multi-national working group. This publication limited itself to the presentation of the appropriate protocols. It did not address the validation of the protocols or provide of quality control limits for control strains. The disc diffusion protocols of Alderman and Smith (2001) have been further developed by a working group set up by the NCCLS that will publish its version in the near future (NCCLS, 2003). This document will present the quality assurance limits for control strains that are suggested as appropriate for some of the disc diffusion assay protocols. The control limits in this document will be provisional and will be published to invite discussion and feedback. Alderman and Smith (2001) did not and the forthcoming NCCLS (2003) document will not report any attempt to address the very difficult issue of establishing interpretive criteria for the results of disc diffusion assays with bacteria associated with fish disease (Smith, 2001).

In any investigation of bacterial susceptibility a choice has to be made as to which method, agar-dilution, broth-dilution, disc diffusion or commercially developed variants of these methods, to employ in any study. In human health studies it is accepted that agar-dilution methods should be treated as the gold standard and data generated by other methods should be compared to the data generated by this gold standard (Piddock, 1990). There is, however, evidence that, in the non-human field, the situation is different. In a report published by the Office International des Epizooties (OIE, 1998) on national surveillance programmes for antimicrobial resistance in livestock production 16 of 19 programmes were reported as using disc diffusion as their primary susceptibility test method. There are little data relating to the relative popularity of the different susceptibility

methods in aquaculture but anecdotal evidence would suggest that, at least in front line diagnostic laboratories, the use of disc diffusion methods is very common.

This present study was undertaken to investigate some of the aspects of the internal validity of the disc diffusion protocols published by Alderman and Smith (2001). Precision, the degree of agreement that can be expected between repeat analyses of the same sample either within one laboratory or between laboratories, probably represents the most important aspect of the internal validation of disc diffusion assays (Smith, 2001). Studies of precision facilitate the establishment of the confidence limits that should be applied to any particular measurement of zone diameters. The size of these confidence limits will, in turn, have implications for the sensitivity that the protocol can be assumed to have and will also have implications for any attempt to related disc diffusion zone data to data generated by other methods. Sensitivity here is defined, following the Welac document (Welac Working Group, 1993), as the smallest difference in zone sizes that can be reliably detected. Clearly if the precision is low, the confidence limits will be relatively large and it will only be possible to be certain that zone sizes detected for two strains represent real differences in their susceptibility if the measurements are numerically widely different. Conversely, small differences in zone size could only be treated as significant if studies had shown that the protocol has a high degree of precision.

Precision studies can also acts as triggers for the initiation of studies of the robustness of a protocol with respect to parameters specified within it. Robustness refers to relationship between variations in any such parameter and the data generated by that protocol. A protocol is said to lack robustness with respect to a specific parameter if small changes in that parameter result in large changes in the data generated. If a protocol possesses a lack of robustness with respect to a particular parameter then any failure, during the drafting of that protocol, to specify the acceptable values of the parameter with sufficient rigour, will result in low precision. Therefore, if the application of a protocol generates data that has unacceptably low precision an investigation of the source of the variation should be initiated. If a low precision can be shown to have derived from variations within a particular parameter then it may be possible to specify the acceptable ranges of this parameter more tightly and thereby improve the precision of the protocol.

In this study both inter- and intra-lab precisions of the disc diffusion protocols outlined in Alderman and Smith (2001) were established in a study involving laboratories in seven countries. Each laboratory was asked, to the best of their ability, to perform a specific series of measurement using the published protocol and a standard control strain. The experimental design also allowed the establishment of the robustness of the protocol with respect to the source of antimicrobial discs and of M6-A compliant Mueller-Hinton agar and with respect to duration of incubation.

2. Materials and methods

2.1. Bacteria

Escherichia coli ATCC 25922 is employed by the NCCLS as the primary control strain for their disc diffusion protocols (NCCLS, 1999b). All disc diffusion assays were

performed using this strain and each laboratory obtained an independent culture of this organism from ATCC.

2.2. Media

All disc diffusion assays were performed using M6-A (NCCLS, 1996) compliant Mueller-Hinton agar (MHA). Each of the seven laboratories in the trial used both a locally purchased (MH1–MH7) and a commonly distributed medium (MH8). MH8 was a Difco product (lot no. 1142004) purchased by and distributed from the Danish laboratory.

2.3. Discs

Assays were performed using the two sets of antimicrobial agent discs shown in Table 1. In the case of flumequine and florfenicol the discs in the two disc sets were not only from the same supplier but were also from the same production lot.

2.4. Disc diffusion protocols

All disc diffusion assays were performed according to the protocols of Alderman and Smith (2001).

2.5. Trial design

The design of the trial followed that outlined in NCCLS M37-A (NCCLS, 1999a). Each laboratory performed 20 independent determinations of the zone size produced by the discs in both disc sets on their locally purchased MHA (MH1–MH7). They also performed 10 independent determinations using both disc sets on the commonly distributed MHA (MH8). An independent assay was defined as one in which an independently grown culture of the test organism *E, coli* ATCC 25922 was used. Only one laboratory (Lab 1) significantly deviated from this design. In its experiments using

Table 1
Antimicrobial agent disc sets used in this study

Agent (amount)		Supplier	
		Disc set 1	Disc set 2
Ampicillin 10 µg	AMP	Oxoid[a]	Mast[b]
Amoxycillin 25 µg	AMX	Oxoid	Mast
Florfenicol 30 µg	FLO	Schering Plough[c]	Schering Plough
Flumequine 30 µg	FLU	Oxoid	Oxoid
Oxytetracycline 30 µg	OTC	Oxoid	Mast
Oxolinic acid 2 µg	OXA	Oxoid	Mast
Sulphamethoxazole/trimethoprim 25 µg	SXT	Oxoid	Mast

[a] Oxoid, Basingstoke, England.
[b] Mast, Liverpool, England.
[c] Schering Plough Animal Health Division, Milan, Italy.

local MHA (MH1), this laboratory included FLO in disc sets 1 but not in disc set 2 and FLU in disc sets 2 but not in disc set 1.

2.6. Statistical analyses

In the analyses of the data generated during this trial the data sets originating from the seven participating laboratories were identified by a randomly assigned number (1–7). For each set and subset of data the range (min–max), the mean ($\overline{X}$), the standard deviation (S.D.) and the percentage coefficient of variation (CV = S.D./$\overline{X}$ × 100) were calculated. In addition the range of zone sizes that would include 95% of the data (95% range) were calculated as 95% range=($\overline{X}$ + 1.96 × S.D.) − ($\overline{X}$ − 1.96 × S.D.). Values of the 95% range were rounded up or down to the nearest whole mm.

Comparisons between multiple data sets were made using ANOVA (SPSS.com) and those between pairs of data by the Student's *t*-test. When the standard deviations were significantly different, Welch's *t*-test (InStat 2.03, Graphpad.com) were employed.

Within any data set generated by one laboratory from one disc on one medium, an outlier was identified as a value that was greater or smaller than the mean of the remaining points plus or minus three times the standard deviation. All outliers were eliminated from that data set before any analysis was undertaken.

2.7. Participating laboratories

The following laboratories participated in the study;

Aquaculture Section, Department of Fisheries and Oceans, Moncton, Canada.
Fish Health Unit, Dept. Primary Industries, Water and Environment, Prospect, Tasmania, Australia.
Fish Disease Laboratory, Danish Institute for Fisheries Research, Frederiksberg, Denmark.
Fish Pathology Laboratory, Istituto Zooprofilattico, Legnaro Italy.
Department of Biological Sciences, California State University, Hayward CA, USA.
Department of Microbiology, National University of Ireland, Galway, Ireland.
Division of Fish Health, Institute of Marine Research, Bergen, Norway.

3. Results and analysis

3.1. Overall summary

In a total of 2900 zone size measurements reported by the seven laboratories only one was identified as an outlier. This measurement, which occurred in the data set generated by Lab 2 from the AMP disc, included in disc set 2 on MH2, was eliminated from any further analysis.

Table 2 presents the summary of the data generated for each agent. In this table the measurements of the zone sizes produced by each agent, made under all media/disc

Table 2
Summary of all data[a]

Parameter	AMP	AMX	FLO	FLU	OTC	OXA	SXT
Mean diameter (mm)	19.5	22.4	26.1	38.4	30.9	31.0	30.7
Standard deviation (mm)	2.6	2.4	3.0	5.6	4.2	3.7	3.4
Coefficient of variation (%)	13.1	10.8	11.6	14.5	13.6	11.8	11.1
Maximum (mm)	31	31	35	52	42	40	39
Minimum (mm)	13	16	17	22	20	22	20
95% range (mm)	14–25	18–27	20–32	28–49	23–39	24–38	24–37
95% range as a percentage of mean	56.4	40.2	46.0	54.7	51.8	45.2	42.3
Number of measurements	420	419[b]	400	400	420	420	420

[a] For each agent the data generated, under all media/disc combinations, by the seven laboratories are analysed as a single data set.

[b] One outlier was eliminated from this data set.

combinations by all seven laboratories, were analysed as a single data set. The data in Table 2 demonstrates that, in the experiments reported here, the protocols of Alderman and Smith (2001) generated data with a very considerable variation. For the seven agents the coefficients of variation (CV) ranged from 11.1% to 14.5%. The mean size of the 95% ranges was 13.7 mm and varied from 9 mm for AMX to 21 mm for FLU. As the 95% ranges will increase as the mean zone sizes increase it is probably more useful to express the 95% ranges as a percentage of their respective means. In this form the values vary from 56% for AMP to 40% for AMX.

3.2. Precision

3.2.1. Intra-laboratory precision

Estimates of the intra-laboratory precision for each laboratory were made by calculating the percentage coefficient of variation of the data they generated. An estimate of intra-laboratory precision could be made by calculating the CV of all the data points produced by each laboratory for each agent. In the design used here each laboratory performed disc diffusion assays for seven agents using four media/disc conditions. An estimate of variance made using all the data would, however, include variance arising from media and disc sources. In their routine performance of these protocols, any particular laboratory would only employ one medium and one disc set. Therefore, the estimates of intra-laboratory precision have been made by calculating the CV of the data sets generated independently for each agent in each laboratory under each of the four media/disc combinations. Table 3 presents the means of these four independently calculated values of CV.

The mean CV values for each laboratory and each agent ranged from 2.2 to 9.4 with an overall average 4.5. There were some significant variations in the estimates of intra-laboratory precision in the seven laboratories. The internal precision obtained by Lab 4 (mean CV 6.2) was, for example, significantly ($p < 0.05$) lower and those of Labs 3 and 6 (mean CV 3.6 and 3.3, respectively) higher than the others. In contrast, when the intra-laboratory precision data was analysed by the individual agent involved, there were no significant differences. However, when all the mean CV values obtained for both penicillin

Table 3
Estimates of intra-laboratory precision

Agent	Mean of the percentage coefficients of variation (CV)[a]							Agent Mean[b]
	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7	
AMP	4.3	7.1	4.9	5.5	9.4	3.3	7.7	5.9
AMX	3.7	6.6	4.3	5.2	5.4	4.0	7	5.2
SXT	3.7	4.0	4.2	6.4	3.7	3.4	5.2	4.4
FLO	5.9	4.2	3.7	6.8	4.5	3.3	4.6[c]	4.7
FLU	4.8	4.7	2.6	7.1	2.2	2.2	3.4[c]	3.9
OTC	4.6	5.5	3.0	7.6	3.5	3.1	5.3	4.7
OXA	5.2	5.5	3.1	4.7	3.9	3.7	4.6	4.4
Lab Mean[b]	4.6	5.4	3.7	6.2	4.7	3.3	5.5	

[a] Values of CV were calculated by treating data points generated under the four different media/disc conditions as separate sets. The values in the table represent, for each laboratory, the means of the four CV values calculated.

[b] Agent Mean represents the mean of the mean CV values presented for each laboratory for that agent and Lab Mean represents the mean of the mean CV values presented for each agent by that laboratory.

[c] Data sets for only three combinations were reported.

agents (AMP and AMX) were compared with those obtained for all other agents, those for the penicillins were found to be significantly higher.

3.2.2. Inter-laboratory precision

Using the common medium, MH8, all laboratories performed 10 assays with both disc sets. These data are, therefore, suitable for making an estimate of inter-laboratory precision which is not influenced by variation in the performance of locally purchased media (MH1–MH7). In order to negate any influence of the disc sets, the data produced by all laboratories with each disc set were analysed separately.

The mean zones sizes recorded by each laboratory when each disc set was used with MH8 were calculated. The CV of these mean zone sizes were calculated for each disc set (Table 4). A comparison of the data in Tables 3 and 4 (Table 6) demonstrates that the mean CV values calculated to estimate inter-laboratory precision were considerably larger than those presented in Table 3 as estimates of intra-laboratory precision. The ratio of the

Table 4
Estimates of inter-laboratory precision

Agent	CV of the means of each of the laboratory data sets on MH8[a]	
	Disc set 1	Disc set 2
AMP	12.8	14.9
AMX	9.9	12.0
FLO	10.0	9.7
FLU	12.8	10.9
OTC	13.1	7.9
OXA	10.1	10.8
SXT	9.8	10.7

[a] Values of CV were calculated from the mean zone sizes of the seven laboratories from the assays which they employed the common MH8 media.

estimates of intra- and inter-laboratory precisions ranged from 1:2.1 to 1:2.7. It is of note that both the mean intra- and mean inter-laboratory precisions were lowest (the CV highest) with respect to AMP (Table 6).

Table 5 presents the result of a rank order analysis of the mean zone sizes reported by each laboratory for each agent on MH8 when the data from both disc sets are treated together. In the rank order analysis the lab reporting the smallest mean zone size for any particular agent was awarded the rank value 1 and that with the largest mean, the rank value 7. Table 5 also shows, in parentheses, the individual laboratory means for each agent as a percentage of the mean of the total data from all laboratories (Table 2). This analysis illuminates one factor that was associated with the low values for the estimates of inter-laboratory precision (Table 4). Some laboratories (notably Lab 2 and 6) consistently reported smaller than average zones for all agents. There were also laboratories (notably Labs 3 and 5) that consistently reported larger than average zones. In the case of Labs 3 and 5 these larger means were particularly notable with respect to the divalent cation influenced agents, FLO, OXA and OTC.

3.2.3. Comparison of intra- and inter-laboratory precisions

Table 6 presents a comparison of the overall estimates of intra- and inter-laboratory precisions. For each agent, the overall estimates of intra-laboratory precision were taken from the right-hand column of Table 3. These figures present the mean value of the mean CVs calculated for the data from each laboratory. The estimates of inter-laboratory precision are calculated as an average of the two data sets presented in Table 4. The data in this table demonstrates that the ratio of CV values representing inter-laboratory precision were, with respect to all agents, over twice the respective values representing intra-laboratory precision.

3.3. Influence of media source on zone size

Each laboratory performed 10 assays for each disc set ($n = 20$) using the common MH8 medium and 20 assays with both disc sets ($n = 40$) on their local media (MH1–MH7). In order to estimate the robustness of the protocol with respect to the source of M6-A compliant MHA, the data generated on the two media types were compared. For each laboratory and each agent the difference between the mean zones on the local media

Table 5
Rank ordering of each lab according to mean zone sizes recorded on MH8

Laboratory	Mean rank order	Rank order of mean zone sizes (individual lab means as percentage of all-lab means)						
		AMP	AMX	FLO	FLU	OTC	OXA	SXT
Lab 6	2.0	3 (95)	3 (97)	4 (96)	1 (88)	1 (90)	1 (91)	1 (83)
Lab 2	2.0	2 (88)	1 (84)	1 (89)	3 (93)	2 (93)	2 (93)	3 (97)
Lab 4	2.4	1 (85)	2 (90)	3 (95)	2 (90)	4 (96)	3 (94)	2 (96)
Lab 1	4.1	4 (97)	5 (105)	2 (93)	4 (94)	5 (98)	4 (94)	5 (100)
Lab 7	5.3	7 (119)	6 (107)	7 (116)	5 (96)	3 (95)	5 (96)	4 (99)
Lab 3	6.0	5 (102)	4 (103)	6 (106)	7 (118)	7 (114)	6 (114)	7 (108)
Lab 5	6.1	6 (110)	7 (110)	5 (105)	6 (116)	6 (113)	7 (116)	6 (107)

Table 6
Summary of intra- and inter-laboratory precisions

Agent	Mean intra-lab precision[a]	Mean inter-lab precision[b]	Ratio
AMP	5.9	13.9	1:2.4
AMX	5.2	11.0	1:2.1
FLO	4.4	9.9	1:2.2
FLU	4.7	11.9	1:2.5
OTC	3.9	10.5	1:2.7
OXA	4.7	10.5	1:2.2
SXT	4.4	10.3	1:2.3

[a] See Table 3.
[b] See Table 4.

(MH1–MH7) and the mean zones on the common media (MH8) were expressed as a percentage of the mean zones on the common media. In order to negate any interaction between the disc source and the media, the percentages were calculated separately for each disc set. The data in Table 7 represents the mean of these two values.

In 28 of the 49 laboratory/agent combinations the use of local media (MH1–MH7) resulted in changes in mean zone size that were less than 5% of the mean zone sizes obtained by that lab on the common media MH8. In 14 cases the effect was between 5% and 10% and in seven cases the effect was over 10%. Of the seven laboratory/agent combinations where the effect of the media was shown to be over 10%, five were reported by Lab 4.

Table 7
Comparison of the influence of media source

Agent	Difference between mean zones on common media (MH8) and local (MH1–MH7) as percentage of the mean zone on common media[a] (frequency of significant media effects indicated in parentheses[b])							Frequency of significant media effects[c]
	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7	
AMP	5.6 (1)	10.1 (2)	4.9 (1)	− 7.2 (2)	4.9 (1)	− 2.7 (1)	4.7 (1)	9/14
AMX	2.8 (1)	2.3 (0)	1.0 (0)	2.3 (0)	2.2 (0)	− 1.2 (0)	4.0 (0)	1/14
FLO[d]	0.9 (0)	− 3.3 (0)	2.6 (0)	16.9 (2)	− 2.7 (1)	3.1 (1)	3.7 (1)	5/13
FLU[d]	5.4 (2)	1.6 (0)	− 0.9 (1)	17.3 (2)	5.4 (2)	2.4 (1)	3.7 (1)	9/13
OTC	9.9 (2)	9.9 (2)	5.0 (2)	20.6 (2)	8.6 (2)	8.3 (2)	7.0 (2)	14/14
OXA	9.9 (2)	7.3 (2)	3.8 (1)	11.1 (2)	8.4 (2)	4.7 (2)	6.6 (2)	13/14
SXT	2.9 (1)	4.4 (2)	0.4 (1)	17.9 (2)	3.2 (1)	− 13.1 (2)	5.2 (2)	11/14

[a] Each number represent the average of the percentages calculated independently from the data generated by the two disc sets. Positive numbers indicate that the mean zones on local media were smaller than those on the common medium. Negative numbers indicate the reverse.

[b] The figure in parentheses refers to the analysis in which the data sets produced by the two disc sets are treated separately. (0) indicates that in neither case were the media effects significant and (1) and (2) indicate that they were significant in one or both cases, respectively.

[c] The frequency of significant media effects in the 14 (or 13) data sets generated for each agent by seven laboratories with both disc sets.

[d] Lab 7 included FLO in disc sets 1 but not in disc set 2 and FLU in disc sets 2 but not disc set 1. All other laboratories included both agents in both disc sets.

For convenience the estimates of media influence shown in Table 7 represent an average of two measurements made on the data generated by each disc set. Statistical analysis is, however, easier to perform when data produced by each laboratory from studies with each disc set are treated separately. The results of this analysis are shown in parentheses in Table 6. The right-hand column of Table 6 also presents the frequency of significant media effects when the two disc sets used are analysed separately. Considering the data from all seven laboratories, significant media effects were reported in 62 of the 98 situations.

The data presented in Table 7 shows clearly that the use of different media had greater effects on the zone sizes produced by some agents than it did with others. For example, all laboratories, in both of their disc sets, reported significant media effects for OTC discs on the two media they employed. In contrast, only one laboratory reported a significant media effect on AMX zones and this was with respect to the data from only one of their disc sets. For each laboratory rank order analysis was used to order the agents from that most susceptible to media variation to that least susceptible. Using the data produced by all seven laboratories this order was OTC>OXA>AMP>SXT>FLU>AMX>FLO.

3.4. Influence of disc source on zone size

With each of the disc sets the laboratories performed 10 assays on the common media MH8 and 20 assays with their local media (MH1–MH7). These data were used to quantify any effect of disc source on the zone sizes. In the case of both FLU and FLO, the discs in disc sets 1 and 2 were not only from the same supplier but were also from the same production batch. Therefore, in the analysis of any disc effects, the data generated from these agents function only as an internal control.

Disc effects were quantified by calculating the difference in the mean zone sizes produced by the discs in the two sets as a percentage of those produced by the discs in set 1. In order to negate any media effects, the data from the common and local media used by any laboratory were analysed separately. The data in Table 8 is an average of these values. Table 8 also presents, in parentheses, the frequency with which statistically significant disc effects were detected.

The most frequently significant and the largest, disc effects were detected with AMP and SXT and the smallest with quinolones, FLU and OXA.

When the data, collected from all labs on MH8, for each agent in each disc set were compared, the disc effects for AMX, FLO, FLU, OTC and OXA were not significant. In the cases of AMP (disc set 1 mean, 20.6 ± 2.6: disc set 2 mean, 19.2 ± 2.9) and SXT (disc set 1 mean, 32.3 ± 3.2: disc set 2 mean, 30.5 ± 3.3) the disc effects were, however, very significant ($p < 0.01$).

3.5. Influence of incubation time on zone size

Two of the six laboratories (Labs 2 and 3) reported zone sizes recorded after 1 and after 2 days of incubation. The data recorded at two times enables an analysis to be made of the robustness of the protocol with respect to the time at which zones are read. For each laboratory the difference between the day 1 and day 2 measurements were determined for

Table 8
Influence of disc source

Agent	Difference between mean zones from disc set 1 and disc set 2 as percentage of the mean zone from disc set 1[a] (frequency of significant disc effects indicated in parentheses[b])							Frequency of significant disc effects[c]
	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7	
AMP	7.0 (2)	− 1.2 (0)	15.4 (2)	5.9 (2)	10.8 (2)	7.2 (2)	1.6 (1)	11/14
AMX	2.4 (1)	0.6 (0)	6.8 (2)	− 0.1 (0)	− 1.4 (0)	6.4 (2)	2.8 (0)	5/14
FLO[d]	6.3 (1)	2.6 (0)	− 0.8 (0)	+2.7 (0)	− 3.0 (1)	+0.8 (0)	+3.3 (0)	2/13
FLU[d]	− 2.9 (0)	+0.6 (0)	+4.7 (1)	− 0.1 (0)	− 1.4 (0)	− 0.7 (0)	+0.5 (0)	1/13
OTC	− 4.2 (1)	− 1.4 (0)	13.1 (2)	− 1.8 (0)	9.8 (2)	6.0 (2)	7.1 (1)	8/14
OXA	− 1.2 (0)	− 0.4 (0)	− 1.1 (0)	− 0.8 (0)	− 2.1 (0)	0.0 (0)	− 2.5 (1)	0/14
SXT	− 4.5 (2)?	− 2.1	9.4 (2)	7.8 (2)	14.2 (2)	10.3 (2)	8.3 (2)	12/14

[a] Percentage difference figures for each laboratory and each agent, represent the average of the values calculated for the data collected on the common MH8 medium and the local MH (1–7) medium. Positive values indicate that the disc set 1 zones were larger than those for disc set 2. Negative values indicate the reverse.

[b] The figure in parentheses refers to the analysis in which the data sets produced by the two media are treated separately. (0) indicates that in neither case were these data sets significantly different and (1) and (2) indicates that they significantly different in one of both cases, respectively.

[c] The frequency of significant media effects in the 14 (or 13) data sets generated for each agent by the seven laboratories from two media.

[d] With respect to Lab 7, data for these agents was only available from MH8.

each of the four media/disc combinations separately. The figures presented in Table 9 represent the mean of these four determinations for each laboratory.

For five of the seven agents the mean zone sizes recorded by both laboratories after 1 day of incubation were larger than those recorded a day later. The greatest changes in zone size over time were those produced by the penicillins (AMP and AMX) and these were significant in 15 of the 16 media/disc conditions. With respect to the two quinolone agents, FLU and OXA, there was a disagreement between the two laboratories as to whether the zones increased as incubation time increased. Lab 3 recorded increases that were significant in three of the four media/disc conditions for FLU and in two media/disc conditions for OXA. In contrast, Lab 2 recorded an overall reduction in the mean zone size for both agents as incubation time increased. These reductions were not significant in any of the four media/disc conditions.

The right-hand column of Table 9 presents an estimate of the inter-laboratory variation in both the day 1 and day 2 data collected from the common medium MH8. The difference between the data of the two labs, as a percentage of the mean for Lab 2, was calculated independently for each disc set at each time of reading; the average of these four calculations are shown in the table. Examination of the data in Table 9 demonstrates that for all agents, the differences in the zone size that are recorded after the two incubation times by either laboratory are smaller than the differences that occur between the two laboratories.

3.6. Influence of incubation temperature on zone size

The Alderman and Smith (2001) protocols specify that incubation should be carried out at 22 ± 2 °C. In these experiments Labs 2–7 reported using an incubation temperature of

Table 9
Comparison of mean zones sizes measured after 1 and 2 days

Agent	Difference between mean zones read on day1 and day 2 as percentage of the mean zone read on day 2[a] (frequency of significant time effects indicated in parentheses[b])		Difference between mean zones from Lab 2 and Lab 3 as percentage of the mean zone from Lab 2 (MH8 data only)[c] (frequency of significant time effects) indicated in parentheses[d]
	Lab 2	Lab 3	
AMP	9.2 (4)	11.2 (4)	12.3 (4)
AMX	8.9 (4)	8.9 (3)	13.7 (4)
FLO	2.5 (2)	1.1 (1)	12.7 (4)
FLU	1.0 (0)	− 3.8 (3)	9.7 (4)
OTC	4.4 (2)	0.7 (2)	15.8 (4)
OXA	1.8 (0)	− 2.9 (2)	15.5 (4)
SXT	5.1 (4)	2.6 (2)	9.7 (4)

[a] Percentage difference figures for each laboratory and each agent in these columns represent the average of the mean values calculated independently from the data sets collected using all four media/disc combinations. Positive values indicate that the mean zones were larger when read on day 1. Negative values indicate the reverse.

[b] The figures in parentheses in these columns indicate the number of the four media/disc combinations in which the data sets recorded at the two times were significantly different.

[c] Percentage difference figures for each laboratory and each agent in this column represent the average of the mean values calculated independently from the data sets collected using both disc sets after both incubation times.

[d] The figures in parentheses in this column indicate the number of the four disc set/reading time combinations in which the data sets recorded by the two laboratories were significantly different.

22 °C but Lab 1 reported using incubation at 24 °C, which is the upper limit of the specified range. A comparison of the data from Lab 1 with that of the other laboratories might provide some information of the robustness of the protocols with respect to temperature of incubation. However, the lack of any intra-laboratory control of temperature as a variable and the overall low inter-laboratory precision (Table 4) place limits on the conclusions that could be drawn from such a comparison. The data presented in Table 5 where all laboratories are ordered on the basis of the relative size of the mean zones they reported for each agent on MH8, shows that the mean rank of Lab 1 was 4.1 (maximum 5; minimum 2), Thus, the data from Lab 1 was never extreme and was, on average, close to the middle of those of the seven laboratories. This limited analysis does not suggest that, within the range 22 °C ± 2 °C, the protocols lack robustness with respect to temperature.

3.7. ANOVA analysis of the relative sources of variance

The experiment reported here was primarily designed to investigate the effect of three variables on the zone sizes that were recorded using the Alderman and Smith (2001) protocols. These variables were as follows: the laboratory performing the analysis, the source of media and the source of the discs employed. F-ratios generated by an ANOVA analysis of the data sets available for each agent were used to generate numerical estimates of the relative importance of these variables (Table 10). Initially this ANOVA analysis was performed on the data from all laboratories. The relative sizes of the F-ratios calculated for AMX, FLO and FLU demonstrated that, with respect to theses agents, inter-laboratory variation alone was the major source of the overall variation. Inter-laboratory variation was

Table 10
Values of F calculated from ANOVA analysis of the total data sets for each agent

Variable	AMP	AMX	FLO	FLU	OTC	OXA	SXT
ANOVA analysis of data from all seven laboratories							
Lab	164	173	172	196	285	301	177
Media	24	12	32	10	441	231	49
Disc set	106	9	na	na	41	<1	86
Media and Lab	9	1	17	24	15	4	43
Media and Disc set	1	7	na	na	2	0.1	52
Lab and Disc set	15	<1	na	na	45	1.5	6
All three variables	2	3	na	na	2	<1	2
ANOVA analysis of data from all laboratories except Lab 4							
Lab	66	45	29	70	37	38	31
Media	2	2	2	<1	11	8	<1
Disc set	3	1	na	na	3	<1	5
All combinations	<2	<1	<1	<2	<2	<1	<3

na indicates that the variable was not applicable.

also demonstrated to be the most significant source of variation for AMP and OXA but, in the case of these two agents, disc source (AMP) and media (OXA) were also shown to be of importance. Of the seven agents, OTC was the only one for which the laboratory as a single variable did not generate the largest F-ratios. For this agent, the analysis suggested that media was the more important variable. The analysis of the SXT data demonstrated a significant degree of interaction between the variables that complicates any interpretation of the F-ratios. Despite this, it is clear that inter-laboratory variation was the most significant source of variation in the total data set.

Examination of the data in Table 7 suggests that the data from Lab 4 showed larger media effects than those of any other laboratory; therefore, the ANOVA analysis was repeated after the data from this laboratory had been eliminated. The F-ratios resulting from the analysis of the data from the remaining six laboratories are also presented in Table 10. This analysis generated a much clearer picture of the relative importance of the variables. The F-ratios generated by the laboratories as a single variable were, with respect to all seven drugs, very much larger than the F-ratios generated by any other variable or combination of variables. With respect to OTC and OXA, the media as a single variable still generated relatively high F-ratios (11.0 and 8.3, respectively).

In summary, the ANOVA analysis demonstrated that, for the data generated by six of the laboratories, inter-laboratory variation was the single most important source of imprecision detected in this work. When the data from all seven laboratories were analysed, the source of media was demonstrated to be a significant variable with respect to the data generated for OTC and OXA.

4. Discussion

The study reported here was undertaken in an attempt to quantify the level of precision that can be expected during the application of the disc diffusion protocols of Alderman and

Smith (2001). The experimental design was adopted in order to facilitate the identification of the major sources of any imprecision. To this end each laboratory was asked, to the best of their ability, to perform a specific series of measurement using the published protocol for disc diffusion assay of antimicrobial agent susceptibility testing.

4.1. Overall precision

The central questions, raised by the degree of variation in total data generated from seven laboratories for each agent (Table 2), are whether this level of precision is acceptable and can it be reduced? There can be no, generally applicable, level of precision that can be regarded as acceptable for all biological assays. Each assay, depending on its inherent properties, will vary in the extent to which it can reasonably be expected to reach any given level of precision. For any particular assay the acceptable level of precision will depend not only on its internal properties but also on the use to which one wishes to put the data it generates. Conversely, the level of precision of any assay will determine the legitimacy of the use of the data the assay generates for any application.

In the data presented here (Table 2), the coefficients of variation for the seven agents examined, varied from 10.8% to 14.5%. The mean size of the 95% ranges was 13.7 mm and varied from 9 mm for AMX to 21 mm for FLU. In other words, if any laboratory used these protocols to generate a zone size for *E. coli* ATCC 25922 against a FLU 30 μg disc, we could, with 95% confidence, expect only that the zone would lie somewhere within a range spanning 21 mm.

One approach to investigating the acceptability of the precision of the protocols studied in this work is by a comparison of the calculated 95% range values (Table 2) with the control limits published by the NCCLS (M31-A) (NCCLS, 1999b) for disc diffusion assays using their protocol and the strain *E. coli* ATCC 25922. As the NCCLS use the 95% range only as a primary guide in setting their control limits (NCCLS, 1999a), this comparison can provide only a general indication of the acceptability of the precision obtained in this work. NCCLS document M31-A (NCCLS, 1999b) presents values for two of the discs used in this study (AMP 10 μg and SXT 25 μg). As the mean zones at the higher temperatures used in the NCCLS protocols result in smaller zone sizes than are produced during the 22 °C incubation used here, it would appear more reasonable to compare the 95% ranges and the control limits as percentages of their mid-points. Using this calculation the NCCLS values for AMP and SXT would be 31.5% and 28.5% and the respective values, calculated from the data collected in this work, would be 50% and 35%. With respect to the 21 discs reported on in NCCLS 31-A, the mean value for the ranges as percentages of their midpoints was 26.3%. The similar figure for the seven discs in this study was 44.6%. These comparisons suggest that the precision obtained employing the Alderman and Smith (2001) protocols is considerably lower that that routinely expected of the NCCLS protocols (NCCLS, 1999b).

4.2. Comparison of the relative importance of intra- and inter-laboratory precisions

Comparison of the estimates of inter and intra-laboratory precision (Table 6) demonstrates that, for each agent, the mean inter-laboratory CV values were over twice the mean

intra-lab values (mean ratio 1:2.4). It should be noted that the estimates of inter-laboratory precision were only obtained from data generated using the common MH8 and that they were calculated independently for the data generated from each disc set. Thus, neither variations in the performance of the locally purchased media nor variation in the discs in the two sets can account for the low estimates of inter-laboratory precision. These comparisons demonstrate that the variation between data generated by the seven laboratories was much greater than the variation in the data generated by any one of them when they employed a single source of discs and a single medium. This conclusion is further strengthened by the results of the ANOVA (Table 10), which demonstrated that, for each agent except OTC, the inter-laboratory variation was the greatest single source of imprecision.

The data presented in this work does not allow any identification of the factor or factors that have contributed to the low inter-laboratory precision. Two suggestions can, however, be made. The first is related to the fact that each laboratory prepared the agar plates they used from commercially supplied powder. Thus, each used their own supply of distilled water. It is possible that at least part of the overall inter-laboratory imprecision may be a function of variations in the quality of the water used by each laboratory. Variation in the divalent cation concentration of the waters might, for example, play some part in explaining the inter-laboratory variation with respect to the OTC and quinolone agents. Cation adjustment procedures have been developed to minimise variations arising from fluctuations in cation concentrations in either batches of MHA or distilled water (NCCLS, 1999b). In essence cation adjustment involves adding Ca^{2+} and Mg^{2+} to MHA until the required results are obtained with control strains. During the discussions that led to the formulation of the Alderman and Smith (2001) protocols, the issue of cation adjustment of MHA was debated but a decision was taken not to include them as a component of the disc diffusion protocols. The data obtained in this work would suggest that this decision should be re-visited.

A second possible explanation for the poor inter-laboratory precision would relate to the reading of zones. Kavanagh (1972), in a study of the theory and practice of disc diffusion assays, concluded that the measuring of zone sizes was the most significant source of variation. The suggestion, that inter-laboratory variation in the measurement of zone sizes plays a major role in reducing inter-laboratory precision, is not inconsistent with any of data presented in this work. If inter-laboratory variations in the reading of zone sizes are an important source of imprecision, then it is possible that greater gains in precision might be expected from efforts to train laboratory personnel than from further refining of the protocols.

4.3. Robustness of the assay protocols

The design of the experiment reported here was based on the assumption that variations in the source of both the media and the antimicrobial agent disc employed had the potential to influence the zone sizes recorded.

4.3.1. Media source

This work provided evidence that variations in the media used were capable, in some situations, of producing statistically significant variations in the zone sizes. One constraint

for all the participating laboratories was that all Mueller-Hinton agar used should be sold as M6-A (NCCLS, 1996) compliant. Thus, these conclusions concerning media effects on zone sizes relevant to variations that occur in agar media made from powders that have passed the quality control procedures laid down in M6-A. In interpreting the data presented in Table 6 it is important to remember that each laboratory employed a different local medium and, therefore, there are no reasons to expect the same media effects from each laboratory. The most notable media effects were those recorded for the five non-penicillin agents tested by Lab 4. With respect to these five agents Lab 4 recorded zone sizes on their local MH4 that were on average 16.8% (range 11–21%) smaller than they recorded on the common MH8. Clearly the performance of discs on MH4 plates was significantly different ($p < 0.05$) to their performance on MH8 plates even when both agars were prepared in the same laboratory.

With respect to all laboratories, statistically significant media effects were both more common and the mean effects larger, with some agents than with others. The two agents showing the greatest and most frequently significant media effects were OTC and OXA. The observation that the zone sizes associated with OTC discs are sensitive to variations in media composition is not new (D'Amato et al., 1975). These effects have been associated with variation in the concentrations of inhibitory divalent cations particularly Mg^{2+} and Ca^{2+}. The activity of the quinolone OXA is also known to be sensitive to inhibition by these ions (Smith and Lewin, 1988), thus, it is not surprising to see that the media effects for both OTC and OXA are similar. If divalent cation variation is the major reason for the media effects seen with OTC and OXA, then it is not immediately clear why the media effects were less dramatic with the other quinolone agent, FLU, which is also sensitive to inhibition by these agents (Pursell et al., 1995). It is interesting to note that in the reports from all laboratories the mean zones for OTC and OXA were smaller on the local media (MH1–MH7) than on the common MH8 media. This suggests, ironically, that it was the MH8, used as a common reference point, which was itself the most deviant with respect to these agents.

4.3.2. Disc source

In analysing the data on the effects of the source of discs on the zone sizes recorded, it must be remembered that a maximum of two suppliers were used for any one agent and that, in each case, discs from a single batch prepared by each supplier were distributed to all laboratories. These constraints limit the generality of any deductions that can be made from the data in Table 7. A further caveat against over-interpretation of these data must be taken from the fact that, on two occasions, statistically significant differences were recorded between the zone sizes recorded for FLO and FLU in disc sets 1 and 2 (Table 7). With respect to these agents, the discs in both disc sets were from the same supplier and the same batch. Thus, these statistically significant differences cannot reflect differences between the discs but must have arisen as a function of other intra-laboratory variables.

With the particular discs used in this study, the most significant disc effects were recorded for AMP and SXT. These observations cannot, however, be used to support the conclusion that similar differences will be observed between all batches of AMP and SXT discs from the two suppliers used in this study. Nor will they support the conclusion that differences in disc performance should be expected to occur more frequently with respect

to these agents than with the others in the study. The only valid conclusion that can be made from the data on disc effects presented in this work is that disc source can be a variable and that the variations in performance between discs from different suppliers may be statistically significant. In this context, it is worth noting that only in three of the 98 laboratory/media situations examined was the difference in the mean zone sizes resulting from variations in disc source greater than 10%.

4.3.3. Incubation time

The protocols of Alderman and Smith (2001) allow reading of zone sizes to be taken at either 22–28 or 44–48 h. The data in Table 8 can provide data as to the robustness of the assay to variations in incubation time. In general there were few consistent differences in the zone sizes recorded for FLO, FLU, OTC and OXA after the two incubation times. With respect to the penicillins (AMP and AMX) and to a lesser extent with SXT the zone sizes recorded after a shorter incubation were significantly larger. These data strongly suggest that the protocols should be modified to specify a more rigorously defined incubation time.

4.3.4. Temperature

The very limited data on the influence of incubation temperature of zone sizes provides no evidence that the $4°$ range (22 ± 2 °C) specified in the protocols requires to be modified.

5. Conclusions

The overall precision of the data generated by the Alderman and Smith (2001) protocols used in this work was disappointing. These protocols were developed as a first step in the formulation of an internationally acceptable standard method for determining the antimicrobial agent susceptibility of bacteria associated with fish diseases. The poor precision must be a cause for concern but it should also act as a stimulus for investigations into where, in the formulation or in the performance of the Alderman and Smith (2001) protocols, the precision has been lost.

This work demonstrates that major source of the low precision was the degree of inter-laboratory variation. The possibility that variations in the cation concentrations of the distilled water used by different laboratories was one source of the inter-laboratory variation would be consistent with the data presented here. Relatively simple experiments could be performed to confirm this hypothesis. If the protocol was shown to lack robustness with respect to distilled water quality, then serious thought should be given to the introduction of cation adjustment procedures into the protocol.

Differences in the measurement of zone sizes by individual scientists also represent a possible source of inter-laboratory variation. Training personnel in the practice of zone reading may be required if this is demonstrated to be a significant source of variation.

There is clear evidence (Table 9) that the length of incubation is an important parameter and that the latitude with which it was specified in the Alderman and Smith (2001) protocols is excessive. It is recommend that the protocols should be amended to allow reading only after 2 days of incubation.

Acknowledgements

References

Alderman, D.J., Smith, P., 2001. Development of draft protocols of standard reference methods for antimicrobial agent susceptibility testing of bacteria associated with fish diseases. Aquaculture 196, 211–243.

D'Amato, R.F., Thornsberry, C., Baker, C.N., Kirven, L.A., 1975. Effect of calcium and magnesium ions on the susceptibility of *Pseudomonas* species to tetracycline, gentamicin polymyxin B, and carbenicillin. Antimicrobial Agents and Chemotherapy 7, 596–600.

Kavanagh, F., 1972. Analytical Microbiology, vol. 2. Academic Press, New York. 253 pp.

OIE, 1998. Office International des Epizooties; Report on Technical Items T 089. Office International des Epizooties, Paris.

NCCLS, 1996. Protocols for Evaluating Dehydrated Mueller-Hinton Agar; Approved Standard. NCCLS Document, vol. M6-A. Wayne, PA, USA. 15 pp.

NCCLS, 1999a. Development of in-vitro susceptibility testing criteria and quality control parameters for veterinary antimicrobial agents. Approved guidelines. NCCLS Document, vol. M37-A. NCCLS, Wayne, PA, USA. 21 pp.

NCCLS, 1999b. Performance standards for antimicrobial disk and dilution susceptibility test for bacteria isolated from animals. Approved guidelines. NCCLS Document, vol. M31-A. NCCLS, Wayne, PA, USA. 80 pp.

NCCLS, 2003. Disk diffusion susceptibility testing protocols for antimicrobials used for aquatic species isolates; A report. NCCLS Document, vol. M42-R. NCCLS, Wayne, PA, USA. 39 pp.

Piddock, L., 1990. Techniques used for the determination of antimicrobial resistance and sensitivity in bacteria. Journal of Applied Bacteriology 68, 307–318.

Pursell, L., Samuelsen, O.B., Smith, P., 1995. Reduction in the in-vitro activity of flumequine against Aeromonas salmonicida in the presence of the concentrations of Mg^{++} and Ca^{++} ions found in sea water. Aquaculture 135, 245–255.

Smith, P., 2001. Accuracy, precision and meaning of antimicrobial susceptibility testing of bacteria associated with fish diseases. Aquaculture 196, 253–266.

Smith, J., Lewin, T., 1988. Chemistry and mechanisms of action of the quinolone antimicrobials. In: Andriole, V.T. (Ed.), The Quinolones. Academic Press, London, pp. 23–82.

Welac Working Group, 1993. Welac Eurochem Guidance Document No. 1. Accreditation for chemical laboratories: guidance on the interpretation of the EN45000 series of standards and ISO/IEC. Guide 25. Addition 1: April.