



HAL
open science

Structured approach to design of diagnostic test evaluation studies for chronic progressive infections in animals

Søren Saxmose Nielsen, Nils Toft, Ian Andrew Gardner

► **To cite this version:**

Søren Saxmose Nielsen, Nils Toft, Ian Andrew Gardner. Structured approach to design of diagnostic test evaluation studies for chronic progressive infections in animals. *Veterinary Microbiology*, 2011, 150 (1-2), pp.115. 10.1016/j.vetmic.2011.01.019 . hal-00687332

HAL Id: hal-00687332

<https://hal.science/hal-00687332>

Submitted on 13 Apr 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

Title: Structured approach to design of diagnostic test evaluation studies for chronic progressive infections in animals

Authors: Søren Saxmose Nielsen, Nils Toft, Ian Andrew Gardner



PII: S0378-1135(11)00038-1
DOI: doi:10.1016/j.vetmic.2011.01.019
Reference: VETMIC 5157

To appear in: *VETMIC*

Received date: 14-10-2010
Revised date: 6-1-2011
Accepted date: 20-1-2011

Please cite this article as: Nielsen, S.S., Toft, N., Gardner, I.A., Structured approach to design of diagnostic test evaluation studies for chronic progressive infections in animals, *Veterinary Microbiology* (2010), doi:10.1016/j.vetmic.2011.01.019

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1 **Structured approach to design of diagnostic test evaluation studies**
2 **for chronic progressive infections in animals**

3
4 *Søren Saxmose Nielsen, Nils Toft, Ian Andrew Gardner*

5
6 ¹*Department of Large Animal Sciences, Faculty of Life Sciences, University of Copenhagen,*

7 *Grønnegårdsvej 8, DK-1870 Frederiksberg C, Denmark*

8 ²*Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California,*

9 *Davis, CA95616, USA*

10 *Corresponding author: Søren S. Nielsen, Department of Large Animal Sciences, Grønnegårdsvej 8,*

11 *DK-1870 Frederiksberg C, Denmark, e-mail: ssn@life.ku.dk, fax: + 4535333022, tel: +4535333096.*

12

13

14 **ABSTRACT**

15 Diagnostic test evaluations (DTEs) for chronic infections are challenging because a protracted
16 incubation period has to be considered in the design of the DTE, and the adverse effects of infection
17 may be widespread and progressive over an animal's entire life. Frequently, the specific purpose of
18 the test is not formally considered when a test is evaluated. Therefore, the result is often a DTE where
19 test sensitivity and specificity estimates are biased, either because of problems with establishing the
20 true infection status or because the test detects another aspect of the infection (and analyte) than
21 originally intended.

22 The objective of this paper is to outline a structured approach to the design and conduct of a
23 DTE for diagnostic tests used for chronic infections in animals, and intended for different purposes.
24 We describe the process from reflections about test purpose and the underlying target condition
25 through considerations of the pathogenesis, and specification of a practical case definition, which can
26 subsequently be used in the DTE for the specific purpose.

27 The process is illustrated by two examples of *Mycobacterium avium* subsp. *paratuberculosis*
28 (MAP) infections in cattle. MAP infections are chronic and can result in different adverse effects at
29 different time points during the incubation period. The description provides input on the process and
30 deductive reasoning which are integral parts to develop a high-quality design of a DTE for chronic
31 infectious diseases.

32

33 **Key words:** chronic infectious diseases; design strategy; diagnostic test evaluation; paratuberculosis

34

35 1. Introduction

36 Chronic progressive infectious diseases such as brucellosis, tuberculosis, leishmaniasis and
37 paratuberculosis pose special challenges in the design and conduct of diagnostic test evaluations
38 (DTEs). Specifically, the long and to some extent unpredictable disease progression can result in large
39 variation in sensitivity and specificity estimates from DTEs of similar tests (Nielsen and Toft, 2008).
40 Lack of explicit focus on the specific utility of a diagnostic test under evaluation, often termed the index
41 test, is a likely explanation, i.e. the purpose of testing has not been made explicit and therefore test
42 results are interpreted in relation to multiple purposes without really being evaluated specifically in
43 relation to these purposes. The World Organisation for Animal Health (OIE) endorses the use of the
44 “fitness for purpose” criterion, where purposes for example could be to demonstrate freedom from
45 infection in a defined population, to confirm a clinical diagnosis, or to determine the immune status of
46 individuals (OIE, 2009). “Fit for purpose” means that test results must be interpreted to a specific
47 meaning and purpose. This purpose needs to be defined *a priori* by the decision makers, who must
48 specify what they want the test to detect (an objective) before a DTE can be initiated. For chronic
49 infections, the timing of the diagnosis, which corresponds to stage of disease, needs to be included in
50 the purpose description.

51 Early diagnosis can for example reduce the risk of transmission of the causative agent and
52 disease progression. However, some animals at risk of progressing to the stage of “disease” or to the
53 stage of “transmission” may never reach these stages, either because they are genetically resistant to
54 developing disease, are protected by vaccination or because they will not live long enough. Therefore,
55 treatment or culling may be a costly alternative to doing nothing. Late-stage disease diagnosis will
56 often result in a high correlation between test result and clinical signs, and thereby increase the
57 decision makers’ confidence in the test-results. However, transmission may already have already
58 occurred without being detected.

59 Many DTEs of chronic infections are flawed by improper designs and inherent biases both from
60 design issues and from lack of consideration of the close relationship between index test – purpose –
61 target and context of application, exemplified by tuberculosis, HIV and malaria in humans (Fontela et
62 al., 2009), and by *Mycobacterium avium* subsp. *paratuberculosis* (MAP) infections in animals (Nielsen
63 and Toft, 2008). We believe that the primary reason for the design flaws stems from not really putting
64 the test and its application into perspective prior to commencing the DTE.

65 The process and general principles for DTEs are described by TDR Diagnostics Evaluation
66 Expert Panel (2006) and Greiner and Gardner (2000a). However, these generic guidelines do not
67 specifically deal with the challenges of chronic infectious diseases in modern veterinary medicine.
68 Thus, the objective of this paper is to provide readers with a structured approach to address the most
69 important issues involved in designing (and conducting) a DTE for chronic, progressive infections.
70 Specifically, we consider biases imposed by the choice of case definition compared to target condition,
71 biases imposed by the choice of design, and biases occurring from the practical issues in conducting
72 the study.

73 To illustrate the specific considerations, we use MAP infection in ruminants (paratuberculosis)
74 as our example. MAP infections cause significant economical losses globally and control is most
75 wanted in countries with a major dairy industry (Kennedy and Nielsen, 2007). Diagnosis is pivotal to
76 infectious disease prevention and control. However, a long and partially unpredictable incubation
77 period makes diagnosis of paratuberculosis challenging. Biased test information inevitably arises,
78 when diagnostic tests are evaluated and used. Specifically, we use two examples of a purpose to
79 illustrate how the initial choices influence the remaining study design decisions.

80

81 **2. Overview of the Diagnostic Test Evaluation (DTE)**

82 A key element in any successful study involving data collection is the planning phase, where the
83 association between the context (including already known features of the problem), the objective of
84 the current study, and the characteristics of the target and study populations are considered. Only
85 when these aspects of the problem and their interrelationships are understood, should planning
86 commence. Thus before considering a study design, many other decisions or choices have to be
87 made. Hence, development of a well-designed DTE includes considerations and understanding of the
88 following elements in the initial planning phase prior to considering a design:

- 89 a) Pathogenesis, including immune responses, predilection sites for the organism, effects of the
90 infection, clinical manifestations and occurrence of any pathognomonic signs;
- 91 b) Analytical principle of the index test, i.e. what analyte does it detect and how is it detected;
- 92 c) Which alternative tests are available, and what are their strengths and limitations, including
93 performance (e.g. sensitivity, specificity and reproducibility) and operational characteristics
94 (e.g. speed of performing the test, ease of sampling, test-costs);

95 d) Context in which the index test is to be used, i.e. what are you trying to identify and why.

96 We believe that the logical flow in planning and designing a DTE should follow Fig. 1. A proper
97 DTE is usually preceded by an initial phase in which an assay is developed and optimised. This phase
98 should not be mixed with the actual DTE. Once the index test has been defined, it should be
99 optimised. The starting point must be consideration of the purpose, i.e. why, how and where will the
100 test be used once it has been evaluated. The optimisation may be closely related to the pathogenesis,
101 which should therefore also be considered in the initial phase. Understanding the purpose and
102 considering the relevant components of the pathogenesis is an iterative process, i.e. if the purpose of
103 the test is to detect animals which may become infectious, then it is important to understand the
104 process leading up to or causing shedding of the infectious agent, whereas a purpose such as
105 confirming clinical disease might make other components of the pathogenesis more important. Only
106 after the purpose has been defined, should the target condition be specified. The target condition is
107 the underlying infection condition that the index test should detect. This condition has to reflect the
108 purpose of testing as much as possible, i.e. if we want to optimise economically-important production
109 measures and control infection, there may be little reason to focus on diseased animals, if infected
110 non-diseased animals can also transmit the agent.

111 After selection of the target condition, one may then proceed to determine the case definition.
112 The case definition is based on the results of the reference standard, against which the index test
113 should be compared. Again, it is imperative that there is a high degree of association between the
114 underlying target condition and the more pragmatic case definition, which may be considered as the
115 practical description of the target condition. The more discordant they are the more problems in terms
116 of biases will be inferred. As an example, the purpose could be to prevent transmission of an
117 infectious agent. The target condition could then be infectious animals (those that currently transmit
118 the agent), and a case definition might be all animals testing positive in a specific agent-detection test.
119 The challenges incurred might be that a test-positive animal excreting low numbers of the infectious
120 agent may not be infectious, or a test-negative animal may become infectious the day after it was
121 tested.

122 Once the purpose, pathogenesis, target condition and case definition have been defined, the
123 design follows logically. There are of course practical considerations and constraints such as time,

124 resources and logistics, which may influence the actual design. However, the process should ensure
125 that the link between purpose and DTE design remains intact as shown in Fig. 1.

126

127 **3. Purpose and target condition**

128 Decision makers should consider a specific purpose for the use of a diagnostic test, for example
129 to maximise profit at the herd-level, to eradicate an infectious agent from a population, or to determine
130 if an animal is infected. With this overall goal in mind, a purpose of testing and a relevant target
131 condition can be selected.

132 The sensitivity and specificity estimates of the index test express the probabilities being test-
133 positive among animals with the condition and being test-negative among animals without the
134 condition, respectively. Specific stages of a chronic infection are different underlying conditions, and
135 the choice of target condition therefore affects both sensitivity and specificity, and ultimately the
136 interpretation of test results. For example, three target conditions for MAP tests were defined in
137 Nielsen and Toft (2008): 1) “infected”, 2) “infectious”, and 3) “affected”, describing three different
138 underlying conditions where the latter two are subsets of the first. Hence, these target conditions are
139 not mutually exclusive, in fact, they should more be seen as successive stages of the disease. It
140 might, however, be possible to move from “infected” to “affected”, without any discernable “infectious”
141 stage. The choice of a specific target condition ultimately relies on the purpose of testing. MAP
142 infected means that the animal is “infected” with MAP; that is MAP have crossed the animal’s natural
143 barriers (usually the intestinal epithelium) and is persisting in the animal without simply being
144 transported passively through the intestinal tract in digested material. This animal could be of
145 relevance for herd classification, or for evaluating the risk of the animal becoming infectious or
146 diseased in the future.

147 The MAP infectious animal sheds sufficient bacteria in milk or faeces to infect susceptible
148 animals horizontally, or can transfer sufficient MAP vertically to result in infection of a susceptible
149 foetus. The sensitivity in this case describes the probability of being test positive if the animal is MAP
150 infectious. The specificity expresses the probability of being test negative among non-infectious
151 animals, i.e. an animal can be infected but non-infectious and still be a non-case.

152 Finally, the MAP affected animal can be of interest because it is adversely affected by the
153 infection, thereby losing financial value because of reduced production performance (milk yield and

154 body condition), although this reduced performance may not be clearly visible to an observer in early
155 phases of infection. For other chronic infections, such as bovine tuberculosis, other purposes and
156 target conditions might be relevant. For example in Denmark, *M. bovis* has been eradicated and the
157 purpose and context of current testing is at the slaughterhouse to demonstrate the continued freedom
158 from infection in the national herd.

159

160 **4. Pathogenesis**

161 Knowledge of the pathogenesis of a chronic infection is a pre-requisite to define a relevant
162 target condition, and thereby the purpose of the study. However, the purpose of the study should
163 usually be the primary motivator for identification of a relevant test, but the purpose can only be
164 specifically described once the pathogenesis is adequately understood. A continuous and iterative
165 update of the purpose - pathogenesis – target condition triad (Fig. 1) is therefore required.

166 For example, MAP infections are presumed to occur in calfhoo (Windsor and Whittington,
167 2010), although time of natural infection is practically impossible to determine in field settings. Adult
168 infections (after first calving) probably occur in some dairy management systems through
169 contaminated feed, drinking water or ingestion of waste water from holding lagoons that is used to
170 flush freestall alleyways. Unpublished data indicates female replacements introduced from a
171 presumed non-infected herd into a MAP infected herd became infected based on ELISA test results
172 and evidence of faecal shedding about 2 to 3 years post entry (Wells et al., 2010). However, it is likely
173 that new adult MAP infections are a small contributor to the total infection pool and for simplicity, a
174 calfhoo infection scenario is the only one considered hereafter.

175 Chronicity of MAP infections is likely one of the most important factors affecting the ability to
176 correctly diagnose “paratuberculosis”, or specifically the defined target condition in a specific study.
177 The host responses to the infection determine the length of the incubation period. However, the host
178 responses to MAP are still poorly understood which makes interpretation of diagnostic test results a
179 challenge. Current understanding of the host responses is described in detail in Coussens (2004).
180 Briefly, pro-inflammatory and anti-inflammatory immune responses both play a role in the
181 pathogenesis of MAP infections. The former may indicate that the animal has control of the infection,
182 whereas the latter may suggest that control of the infection is lost (Coussens, 2004). Pro-inflammatory
183 responses include production of cytokines such as interferon- γ and interleukin 10 (IL-10; Nagata et al.,

184 2010) and IgG2 type antibodies, whereas anti-inflammatory responses usually are characterized by
185 antibodies of the IgG1 type (Koets et al., 2001).

186 The transition from pro- to anti-inflammatory responses is unpredictable, although animals in the
187 range from 2-5 years of age are more likely to be shedding MAP and have IgG1 antibodies than older
188 and younger animals (Nielsen and Ersbøll, 2006). Whether anti-inflammatory responses can revert
189 back to pro-inflammatory responses is not known. The anti-inflammatory phase is usually
190 accompanied with bacterial shedding (Collins, 2002; Nielsen, 2008), declines in milk production
191 (Nielsen et al., 2009) and reduction in body weight (Kudahl and Nielsen, 2009). Furthermore, end-
192 stage infection can be characterised by extensive bacteraemia (Alonso-Hearn et al., 2009), and the
193 animal can die from the infection if not slaughtered. A summary of the current understanding of the
194 pathogenesis and the effects of MAP infections is shown in Fig. 2.

195

196 5. Case definition

197 Target conditions are underlying entities, which we should define in our DTE. The practical
198 realisation of our target condition is the case definition which can be made both at both the animal and
199 herd level. The case definition must reflect the target condition. If the target condition and case
200 definition differ, misclassification bias occurs.

201 A clear and rule-based definition of a case is required in most situations. All non-cases can then
202 be considered controls, or some non-cases can be excluded. Still, exclusion should be considered
203 carefully. For example, if a case is a cow infected with agent X and presenting with diarrhoea, then all
204 other cows are non-cases. However, diarrhoea can be caused by both agent X and agent Y. If the
205 diarrhoea is actually caused by agent X, misclassification of the cases occur, but exclusion of the cow
206 can result in spectrum bias (see Section 6.2 for details). One or more diagnostic reference tests can
207 be used to define the case definition, but these tests should be independent of the index test. For
208 example, if a case definition of “affected by agent X” is specified by occurrence of diarrhoea and
209 antibodies to agent X, then we should not evaluate an antibody test against this case definition. Once
210 we have defined a case definition, we can start considering relevant index tests. In some situations, a
211 case is not defined, *a priori* and latent class statistical methods are used to analyse data. The case
212 definition in these situations is dependent on the joint results of the diagnostic tests, and a “latent”
213 undefinable condition constitutes our case definition. For example, if two antibody tests are used and

214 evaluated against each other, the latent condition is presence of antibodies rather than infection
215 status. If an agent-detection and an antibody-detection test are used, the latent condition reflects the
216 mutual condition that exists between the presence of pathogen (infection) and immune-response
217 (response to infection). These undefinable conditions may be difficult to interpret, particularly if no prior
218 knowledge exists about how the co-existence of the two analytes relates to pathogenesis.

219

220 **6. Bias**

221 Biases are systematic errors, which result in under- or overestimation of the parameters of
222 interest, here primarily sensitivity or specificity. Complete random sampling as part of the study design
223 should result in unbiased estimates, whereas bias resulting from host responses may still be present
224 and should be taken into account when assessing generalisability of results.

225

226 6.1. Bias resulting from host responses

227 False-negative and false-positive reactions are highly likely to occur with most tests, resulting in
228 reduced test sensitivity and specificity, respectively. If infection has not reached a specific disease
229 stage, analytes characteristic of this stage cannot be expected and thus will not be detectable. If
230 infection has reached one disease stage and reverted (or been cleared), the analytes may still be
231 present. Furthermore, the virulence of strains of a pathogen may differ and affect the subsequent
232 pathogenesis and immune response.

233

234 6.2. Bias resulting from study design

235 Biases are hard to avoid in DTEs of chronic infections, primarily because multiple infection
236 markers (e.g. antibodies) occur relatively late. These misclassifications resulting from the sequence of
237 biological processes are usually what we want to elucidate, i.e., we want to demonstrate how the
238 index test deviates from the target condition. Estimates of sensitivity and specificity can be used to
239 provide more reliable estimates of prevalence (Greiner and Gardner, 2000b), if the sensitivity and
240 specificity estimates are unbiased. However, biases in sensitivity and specificity estimates resulting
241 from design of the DTE should be avoided.

242 These design-type biases include different types of selection bias such as “disease progression
243 bias”, “diagnostic access bias”, “work-up bias” or “verification bias”, “incorporation bias”, “loss to follow-

244 up bias”, and “spectrum bias”. Other biases can occur in DTEs in general (Kelly et al., 1997), but the
245 above-mentioned frequently occur in veterinary medical studies, and many other types are special
246 cases of selection bias.

- 247 a) Disease progression bias can occur, if the time between sampling for the DTE and the
248 reference test is so long that disease has progressed from one stage to another. A biological
249 plausible time-span must be selected to avoid this type of bias, e.g. for a chronic infection like
250 MAP, little disease progression is expected to occur in one week, whereas progression from
251 one stage to another can occur in 3 months. Furthermore, if animals are intermittently
252 excreting a pathogen, and the pathogen is present on one test-day but not the next, infection
253 can have progressed/regressed and therefore the case definition may vary between test-days.
254 One solution for dealing with intermittent excretion is to have a case definition aggregated
255 over multiple testing days, e.g. pathogen was excreted at least once during a week.
- 256 b) Diagnostic access bias can occur if part of the study population is not geographically
257 representative of the target population, e.g. a negative reference population is selected from
258 an area considered to be free of the infection. If this area is also free of other infections, which
259 may induce cross-reactions in the index test, diagnostic access bias occurs. In this particular
260 example, we would overestimate the specificity.
- 261 c) Work-up bias or verification bias occurs if the target condition is more likely to be verified for
262 some study objects than for others. Assuming that animals subject to verification are a random
263 sample of the index test positive and index test-negative groups and the same verification
264 procedure is used, then the bias is correctable statistically (Begg and Greenes, 1983).
265 However, verification bias correction may be inadequate when the sensitivity of the screening
266 test is close to perfect (Cronin and Vickers, 2008). In contrast, if a cow with clinical disease is
267 subject to several reference tests to verify that infection is present, whereas the non-clinical
268 cow is subject to only one reference test, the induced verification bias is not correctable,
269 unless at least some of the non-cases are verified. Blinding of results of the index test and
270 reference standard to those performing the test should prevent many forms of verification bias.
- 271 d) Incorporation bias occurs when the definition of the target condition is partially based on the
272 index test. If one antibody and four bacteriological tests are carried out on samples from the
273 same animal, and the case definition is that at least one should be positive in one of the

- 274 bacteriological tests for a sample to be considered a case, then the bacteriological tests
275 should not be included as index tests, because the bacteriological tests are part of the case
276 definition.
- 277 e) Loss to follow-up bias occurs when study subjects are more likely to exit a study because of
278 positive compared with negative test results.
- 279 f) Spectrum bias (Ransohoff and Feinstein, 1978) is perhaps the most frequent type of bias in
280 DTEs of chronic infections. This type of bias occurs when animals in specific infection stages
281 are more likely to be sampled than a randomly-selected animal from the target population.
282 Furthermore, many DTEs are carried out in high-prevalence herds because these herds have
283 more diseased animals and therefore, fewer herds need to be visited to obtain the required
284 sample size. However, the spectrum of disease is often different in high-prevalence herds
285 compared to low-prevalence herds, for example because they are in a different temporal
286 phase of an epidemic. Disease spectrum in a herd could also be affected by interventions,
287 where the most diseased or infectious animals are removed because of positive test results.

288

289 Regardless of its underlying source, bias in individual animal-level sensitivity and specificity
290 estimates can result in substantial bias in herd sensitivity and herd specificity estimates when the latter
291 are obtained by extrapolation of individual values to “herd tests” of varying size (Christensen and
292 Gardner, 2000).

293 To avoid bias associated with design issues, the above should be considered carefully in the
294 design phase. The QUADAS tool (Whiting et al., 2003) can be used as a check-list to avoid these, but
295 considerations of sampling and case definition are key factors to avoid bias. The best way of avoiding
296 these biases is use of complete random sampling of the animals (and herds) used in the study. A
297 biased sampling frame such as selection of cases and controls from a population of animals visiting a
298 veterinary clinic may occur because these animals do not necessarily represent the target population,
299 For example, young animals are often at a veterinary clinic for vaccination purposes and may be
300 selected as controls, but they are less likely to have a chronic disease because they are young.

301

302 **7. Study designs**

303 Many study designs can be used for DTEs. Essentially, the classic observational studies (cross-
304 sectional, case-control or cohort types), or derivatives thereof, can be used. An overview is given in
305 Table 1, although other derivatives also can be used. The choice of study design depends on the
306 choices and effects outlined in previous sections and shown in Fig. 2.

307 Choice of study design often depends on practical issues such costs and logistics. Detailed
308 pathological examination may not be feasible in some countries and regions, whereas a high-
309 throughput sampling system based on milk samples is easy to implement. Costs of testing may
310 determine, if a study can be funded. However, these testing costs may vary considerably among
311 regions because of variable experience with the assay and frequency of use, and different labour and
312 equipment costs.

313

314 **8. Examples**

315 Two examples are provided in the following sections and the implications of decisions made in the
316 process are discussed.

317

318 *8.1. Example 1. Eradication of MAP*

319 *8.1.1. Purpose and target condition*

320 Animal health officials may have decided that MAP needs to be eradicated from a defined
321 population such as a herd or a region. To achieve this goal, a diagnostic test is needed to detect all
322 infected herds and animals with removal of test-positive animals. Given this purpose, an objective for
323 the planned DTE could be to evaluate a diagnostic test, which can be used to detect MAP infected
324 animals, because they pose a risk for future spread of MAP. Our target population is given by the
325 definition of the population, i.e. all animals in the specified population that can be or are infected with
326 MAP.

327

328 *8.1.2. Pathogenesis*

329 The purpose “eradication” requires that we focus on detection of “MAP infected” animals as our
330 target condition. Multiple questions should then be considered in relation to the pathogenesis: How do
331 we define an infected animal? Can infected animals clear the infection? Which test is most likely to
332 detect the infected animal, i.e. which analytes (specific cytokines, antibodies, MAP etc) have the

333 highest probability of being present. If the animal is truly infected, what will be the most cost-effective
334 action to take to achieve the purpose? Currently, we do not know if MAP infections can be cleared by
335 the cell-mediated immune reactions. If not, early detection of infection should target cell-mediated
336 immune reactions, if specific targets are available. MAP specific interferon-gamma detection has been
337 proposed in animals >15 months of age, as younger animals may have non-specific circulating
338 interferon-gamma (Jungersen et al., 2002; Olsen et al., 2005). Detection of interferon-gamma or other
339 cytokines from the cell-mediated immune responses may therefore be appealing because of the
340 potential early detection, but has the disadvantage that non-specific analytes may be present without
341 infection ever being present, and infection may have been cleared. A robust cell-mediated immune-
342 response may contain the infection for the remainder of the animal's life, and a farmer may be
343 reluctant to cull this cow, as the decision may not be cost-effective. However, such reasoning must be
344 deemed irrelevant considering the choice of target condition. If a cost-effective diagnosis for this
345 particular farmer's purpose is required, we might have to change the target condition.

346 Detection of MAP specific IgG1 antibodies produced during anti-inflammatory immune reactions
347 may also be appealing. A disadvantage is that the infection should have progressed significantly, and
348 animals are usually 2 to 5 years of age, or even older, before infection can be detected. An advantage
349 is that infection may have almost have progressed to the stage, just before the animal starts shedding
350 infectious doses of MAP. If animals are tested regularly, they may be detected before this occurs.

351 Furthermore, logistics about sampling of milk or serum to detect IgG1 antibodies is currently
352 much easier than collection of samples for measurement of interferon-gamma, because the latter
353 method requires live T-cells in the samples when they arrive at the laboratory. The last option
354 mentioned here is detection of MAP through culture or PCR techniques. During pro-inflammatory
355 immune responses, the MAP or MAP DNA may be shed in insufficient amounts to be detected, and
356 when shed in significant amounts, the detection may be too late. Considerations about possible false-
357 positive and false-negative test-results should be made for the agent-detection tests.

358

359 8.1.3. *Case definition*

360 We wish to detect the MAP infected animal, but have to find an appropriate reference method to
361 define this in practice. Previous research suggests that MAP can be present in the animal without
362 being detected even if the predilection sites in the intestinal tissues are sampled. Samples from 100

363 tissues from multiple sites are suggested for culture (Whitlock et al., 1996). However, obtaining this
364 number of tissues may be practically infeasible and too costly to motivate if a large sample size is
365 required to obtain required precise sensitivity and specificity estimates. A feasible sample might
366 include 5 intestinal tissues (e.g. 2 x jejunum, 1 x ileo-cecal valve, 1 x colon and 1 x cecum), and 5
367 lymph nodes (e.g. 1 x retro-pharyngeal lymph nodes (Lnn.), 1 x hepatic Lnn., 1 x jejunal Lnn., 1 x ileo-
368 cecal Lnn., 1 mesenteric Lnn.). Culturing these 10 samples would yield 10 results, and it should be
369 specified how to aggregate them into an overall result for the animal. For example, if only a single
370 sample is positive, the animal could be considered a case, and all other animals would be considered
371 non-cases.

372

373 8.1.4. *Bias from host responses*

374 Sensitivities of tests detecting anti-inflammatory responses and bacterial shedding to detect the
375 target condition of MAP infected are likely to be highly affected by chronicity of infection, although age
376 can to some extent be used to improve the diagnostic predictions.

377 The specificity of tests detecting pro-inflammatory responses can be difficult to interpret,
378 because it is not known if infected animals can eliminate the infection (Mikkelsen et al., 2009). If
379 infections can be eliminated, and persisting indicators of pro-inflammatory responses are present,
380 these will result in false-positive reactions after the elimination.

381

382 8.1.5. *Bias from study design*

383 Choice of the case and non-case definitions can significantly affect results. A typical situation
384 with spectrum bias from studies of the target condition of MAP infected could be use of banked
385 samples. If a case was defined as an animal with detectable excretion of MAP in 3 g of faeces, and
386 the target condition was "infected with MAP", all infected animals should be expected to excrete
387 detectable amounts of MAP in 3 g faeces, which is unlikely in early stages of infection. Choosing all
388 cases among the animals with significant shedding of MAP would therefore impact the distribution of
389 animals towards those with more severe infection with resultant higher test sensitivity values.

390 For specificity studies, it is not uncommon to define as non-infected, herds that have previously
391 been found negative by herd screening using an antibody ELISA. The lack of test-positive reactions
392 may be an indicator of lack of infections caused by MAP-related mycobacteria such as *M. avium*

393 subsp. *avium* in this herd. If these closely-related mycobacteria are common in other herds of the
394 target population, diagnostic access bias is introduced, and the specificity might be overestimated.
395 Furthermore, if the antibody ELISA is subsequently evaluated in this population, incorporation bias is
396 introduced, because the antibody ELISA is part of the case definition: the non-infected animal is
397 defined as an animal derived from a previously ELISA-negative herd.

398

399 8.1.6. *Sampling*

400 The sample must reflect the population, which is the general dairy population. Thus some
401 element of random sampling is required. The challenge is most likely that the case definition (and not
402 least the conclusive classification of a non-case) will require some choices with respect to targeted
403 sampling. For example, selecting cases might be based on prior knowledge of MAP being confirmed
404 in certain herds and controls might be selected from (closed) herds without history of disease. Thus, at
405 herd level the sampling might be considered purposive to ensure that both cases and controls can be
406 established. However at the individual animal level, random sampling should be applied to avoid
407 further biases.

408

409 8.1.7. *Study design*

410 A random cross-sectional sample of animals would be ideal to avoid spectrum bias. The desired
411 precision of the sensitivity and specificity estimates should be determined from sample size
412 calculations (Jacobson, 1998). The study animals should be subject to pathological examination
413 immediately after the sample for the index test has been obtained. Sampling of animals at a
414 slaughterhouse may result in spectrum bias, because these animals may be more likely to have been
415 culled because of late-stage MAP infection, or simply because they are old, and older animals are
416 more likely to have late-stage infection than younger animals. On the other hand, the most advanced
417 cases i.e. cows dying from infection and those euthanized at the herd are not included in the sample.

418 An alternative is to obtain a random sample of the animals and conduct a longitudinal study
419 using a reference test and the index test at slaughter to classify the animals. The advantage of this
420 method is that exit status of the animal would show if the animals have actually progressed to the
421 infection stage of MAP infectious, in which the animals transmit MAP. A disadvantage is that the study
422 period can be quite long because some animals initially subjected to the index test may live for many

423 years before they exit the herd. Furthermore, the actual infection status will not necessarily be
424 determined. Therefore, the actual target condition should be redefined to “a MAP infected animal,
425 which does not reach the MAP infectious stage in a life-time”. Animals removed from the herd,
426 regardless of whether for MAP infection or other reasons, would be expected to do this at the same
427 rate as in the target population, and therefore reflect this adequately. Furthermore, if the animals are
428 tested repeatedly, and they are test-positive early but become test-negative later, these animals may
429 be considered false-positive and therefore the specificity can be assessed. However, a cow that is
430 test-positive on the last test before she is slaughtered cannot be used in the study, because we would
431 not know if she is a true positive or a false-positive.

432

433 *8.2. Example 2: Confirmation of clinical paratuberculosis*

434 *8.2.1 Purpose and target condition*

435 A narrower purpose than described in Example 8.1 is confirmation of MAP infection in a clinical
436 suspect animal. This purpose is relevant to veterinarians and farmers, who suspect that a clinically
437 diseased animal may be infected with MAP, and that MAP is the cause of the disease observed. The
438 target population is now only diseased animals, but there is a need to refine this further, e.g. an animal
439 with pneumonia would normally not be included, because MAP is not known to cause pneumonia (see
440 section 4 on Pathogenesis). The target population could be described as “cows with chronic or
441 intermittent diarrhoea”, or “cows with milk production significantly below herd-mates in the same parity
442 and stage of lactation”. This cow is affected by a disease causing entity, whether this is infectious or
443 not. Our purpose is to detect MAP infection in clinically-affected animals, and the target condition is
444 therefore the “MAP affected animal”. Notice that “MAP affected” and “clinical diseased” may differ, in
445 that “clinical” may imply that we have observed that the cow is diseased, but MAP affected is really an
446 underlying condition. However, observational skills may be imperfect and therefore we are not capable
447 of determining if the animal is in fact diseased or not.

448

449 *8.2.2 Pathogenesis*

450 MAP infection may lead to diarrhoea, weight loss and decreased milk production. However,
451 multiple conditions may result in these effects. The description of the target condition is therefore of
452 paramount importance: Are all effects relevant, or just one? If MAP are causing this condition, which

453 analytes are likely to be correlated to disease progression? Production of IgG1 antibodies and
454 excretion of large amounts of MAP have both been associated with late-stage infection. However, both
455 may also be detectable at earlier stages. Furthermore, in the end-stage of infection, the animal may
456 become anergic with no detectable antibodies. However, antibody and agent detection tests may still
457 be highly relevant, and logistics and costs may be the next items to consider in the choice of test. We
458 would have to go back and refine our purpose to determine, if we want to confirm the diagnosis most
459 accurately, and therefore would have to compare the two tests, or if we want the earliest diagnosis to
460 be able to take action on the clinical suspect.

461

462 8.2.3. Case definition

463 A case definition for MAP associated clinical disease could be straightforward: 1) chronic
464 diarrhoea for > 10 days, or 2) a deviation in the milk production of more than 20% of herd-mates in the
465 same lactation and stage of lactation, or 3) weight loss, should be present along with detection of MAP
466 in faeces. While we would most often expect both excretion of detectable amounts of MAP and
467 occurrence of antibodies in the clinical suspect, both definitions could be appropriate, although we
468 could risk that there are no MAP in the specific part of the faeces collected, or the animal is in the
469 anergic phase without production of antibodies. Notice that for example the cow with diarrhoea > 10
470 days but with no detection of MAP in faeces is a non-case and should therefore be included in the
471 specificity estimation.

472

473 8.2.4. Bias from host responses

474 Detection of the target condition MAP affected is probably the MAP condition least affected by
475 chronicity, because the occurrence of IgG1 antibodies as well as detection of bacterial shedding in
476 faeces is well correlated with the occurrence of clinical disease. However, antibodies to MAP can
477 occur prior to the animal's detectable shedding of MAP and occurrence of clinical signs (Nielsen,
478 2008; Nielsen et al., 2009). Therefore, "false-positive" test results prior to occurrence of the target
479 condition MAP affected are highly likely to occur, simply because the target condition has yet to occur.
480 False-positive reactions to any immuno-diagnostic test may also occur if antigens in the immuno-
481 assay bind to cross-reacting antibodies from related organisms.

482

483 8.2.5 *Bias from study design*

484 Defining a MAP clinical suspect may seem trivial, but because the three clinical signs “milk
485 production deviation”, “chronic diarrhoea” and “weight loss” are not specific to MAP infection, we will
486 likely want to include another parameter such as detection of MAP in faeces in the case definition. In
487 this situation we cannot evaluate MAP detection tests without introducing incorporation bias.

488

489 8.2.6. *Sampling*

490 Diagnostic laboratories receiving samples to confirm the diagnosis may be an excellent choice
491 to use as a study population. However, it should be noted that these animals may not all meet the
492 case definition. If only samples from cows with clinical diarrhoea are submitted to the laboratory, cows
493 with deviating milk yield are not part of the case definition, thereby introducing spectrum bias.
494 Therefore, either the purpose and target condition should be changed, or a different study population
495 is needed. Hence, at this stage we should reconsider the target condition. Reduced milk production is
496 a non-specific indicator of a health problem and in some systems this is not really a measurable
497 condition except when the monthly milk testing is carried out. Thus, we should consider prior to the
498 study if it will be possible to define in practice, or if the case definition should only contain the
499 intermittent weight loss without loss of appetite and diarrhoea.

500

501 8.2.7. *Study design*

502 A case-control study can be used and will likely be the most cost-effective and will not
503 necessarily include bias. The most important consideration is the case definition, and controls should
504 include all animals not fulfilling the case definition to avoid spectrum bias.

505

506 **9. Discussion**

507 Inferences about the performance characteristics of diagnostic test results is the ultimate goal of
508 any DTE study, and reliable estimates of sensitivity and specificity are pivotal to reduce
509 misinterpretation attributable to misclassification bias. The structured approach described herein is
510 intended to aid in the planning of a DTE so that minimally biased estimates of sensitivity and
511 specificity for different target conditions may be obtained.

512 First, a purpose and a target condition must be defined, and subsequently a case definition
513 must be established to classify test subjects according to the target condition with minimal errors. The
514 most appropriate and cost-effective study design depends on the relevant target condition, the
515 possibilities for creating a useful case definition and the logistical requirements for carrying out the
516 study. A single study design cannot be deemed the only one for a specific purpose, but clearly some
517 designs are more likely than others to provide biased estimates.

518 We have discussed two examples, in which some choices have been made. The two chosen
519 target conditions “MAP infected” and “MAP affected”, and their corresponding target populations “the
520 general dairy population” and “cattle with enteric diseases” reflect the situations about which we can
521 make inferences about results when the index test we evaluate is to be used in practice. Use of the
522 results from the test evaluation is only applicable to the target condition and target population for
523 which they have been evaluated. Extrapolation to other target conditions is generally unwarranted, but
524 extrapolation to other similar target populations may in some instances be reasonable, although tests
525 should usually be evaluated in the target population. The tests cannot be used for the ‘other’ purpose,
526 but may easily be transferred to other similar populations.

527 Specifying case definitions is also challenging, and it can be difficult to justify some of the
528 choices. Why should 10 tissues be selected rather than 12 tissues? Can 20% deviation in milk
529 production be measured under practical circumstances, and are 15 or 20% more appropriate cut-
530 values? This can be a very practical issue, and the best advice would probably be to replicate
531 definitions used by others to facilitate among-study comparisons. The choices made throughout the
532 process affect the subsequent steps, and can introduce bias. However, the effects imposed by the
533 choices must be considered to differentiate between target conditions, between case definition etc.
534 and therefore it is not possible to provide prescriptive guidelines, which covers all possible scenarios
535 and diseases. We have refrained from providing an “ideal study design”, because we believe that this
536 study design may not really exist, or is practically and logistically impossible to achieve. However, this
537 does not mean that all designs are useful.

538 Equally important to the design of the test-evaluation is the reporting (Gardner, 2010). All
539 elements discussed above should be mentioned. Detailed guidelines for reporting are proposed in
540 Gardner et al. (2011).

541 The approach described here is given on a starting point where decision makers present a
542 problem, for which a diagnostic test should be developed and evaluated. An alternative situation could
543 occur where a new test has been developed, and the goal is to determine for which purpose(s) the
544 new test can be used. In principle, the same considerations apply and only the starting point in Fig. 1
545 is different. However, because Fig. 1 is a triad, the starting point does not impact the reflection steps
546 required.

547 Throughout the DTE, it is important to continuously consider the purpose and target condition,
548 the pathogenesis and case definition. They are closely interrelated and refining them may be required
549 throughout the process to design the ideal DTE for the relevant situation. Ultimately, the proper use
550 and interpretation of the results of the DTE defines the success of the DTE. If the goal is to compare
551 sensitivities and specificities among tests (and determine which tests are superior in one or the and
552 not to establish absolutes for each, then the effect of some biases may be reduced. This would
553 necessitate a head-to-head comparison of multiple tests on the same set of samples

554

555 **References**

- 556 Alonso-Hearn, M., Molina, E., Geijo, M., Vazquez, P., Sevilla, I., Garrido, J.M., Juste, R.A., 2009.
557 Isolation of *Mycobacterium avium* subsp. *paratuberculosis* from muscle tissue of naturally
558 infected cattle. *Foodborne Pathog. Dis.* 6, 513-518.
- 559 Begg, C.B., Greenes, R.A., 1983. Assessment of diagnostic tests when disease verification is subject
560 to selection bias. *Biometrics.* 39, 207-215.
- 561 Collins, M.T., 2002. Interpretation of a commercial bovine paratuberculosis enzyme-linked
562 immunosorbent assay by using likelihood ratios. *Clin. Diagn. Lab. Immunol.* 9, 1367-1371.
- 563 Christensen, J., Gardner, I.A., 2000. Herd-level interpretation of test results for epidemiologic studies
564 of animal diseases. *Prev. Vet. Med.* 45, 83-106.
- 565 Coussens, P.M., 2004. Model for immune responses to *Mycobacterium avium* subspecies
566 *paratuberculosis* in cattle. *Infect. Immun.* 72, 3089-3096.
- 567 Cronin, A.M., Vickers, A.J., 2008. Statistical methods to correct for verification bias in diagnostic
568 studies are inadequate when there are few false negatives: a simulation study. *BMC Med. Res.*
569 *Methodol.* 8, 75.

- 570 Enøe, C., Georgiadis, M.P., Johnson, W.O., 2000. Estimation of sensitivity and specificity of diagnostic
571 tests and disease prevalence when the true disease state is unknown. *Prev. Vet. Med.* 45, 61-
572 81.
- 573 Fontela, P.S., Pant Pai, N., Schiller, I., Dendukuri, N., Ramsay, A., Pai, M., 2009. Quality and reporting
574 of diagnostic accuracy studies in TB, HIV and malaria: evaluation using QUADAS and STARD
575 standards. *PLoS One.* 4, e7753.
- 576 Gardner, I.A., 2010. Quality standards are needed for reporting of test accuracy studies for animal
577 diseases. *Prev. Vet. Med.* 97, 136-143.
- 578 Gardner, I.A., Nielsen, S.S., Whittington, R.J., Collins, M.T., Bakker, D., Harris, B., Sreevatsan, S.,
579 Lombard, J.E., Sweeney, R., Smith, D.R., Gavalchin, J., Eda, S., 2011. Consensus-based
580 reporting standards for test accuracy studies of paratuberculosis in ruminants. Submitted to
581 *Prev. Vet. Med.*
- 582 Greiner, M., Gardner, I.A., 2000a. Epidemiologic issues in the validation of veterinary diagnostic tests.
583 *Prev. Vet. Med.* 45, 3-22.
- 584 Greiner, M., Gardner, I.A., 2000b. Application of diagnostic tests in veterinary epidemiologic studies.
585 *Prev. Vet. Med.* 45, 43-59.
- 586 Jacobson, R.H., 1998. Validation of serological assays for diagnosis of infectious diseases. *Rev. sci.*
587 *tech. Off. int. Epiz.* 17, 469-486.
- 588 Jungersen, G., Huda, A., Hansen, J.J., Lind, P., 2002. Interpretation of the gamma interferon test for
589 diagnosis of subclinical paratuberculosis in cattle. *Clin. Diagn. Lab. Immunol.* 9, 453-460.
- 590 Kelly, S., Berry, E., Roderick, P., Harris, K.M., Cullingworth, J., Gathercole, L., Hutton, J., Smith, M.A.,
591 1997. The identification of bias in studies of the diagnostic performance of imaging modalities.
592 *Br. J. Radiol.* 70, 1028-1035.
- 593 Kennedy, D.J., Nielsen, S.S., 2007. Report from the first IDF ParaTB Forum. Proceedings of the 1st
594 ParaTB Forum, Shanghai, China, Oct. 19, 2006. *Bulletin of International Dairy Federation.* 410,
595 3-7.
- 596 Koets, A.P., Rutten, V.P., de Boer, M., Bakker, D., Valentin-Weigand, P., van Eden, W., 2001.
597 Differential changes in heat shock protein-, lipoarabinomannan-, and purified protein derivative-
598 specific immunoglobulin G1 and G2 isotype responses during bovine *Mycobacterium avium*
599 subsp. *paratuberculosis* infection. *Infect. Immun.* 69, 1492-1498.

- 600 Kudahl, A.B., Nielsen, S.S., 2009. Effect of paratuberculosis on slaughter weight and slaughter value
601 of dairy cows. *J. Dairy Sci.* 92, 4340-4346.
- 602 Mikkelsen, H., Jungersen, G., Nielsen, S.S., 2009. Association between milk antibody and interferon-
603 gamma responses in cattle from *Mycobacterium avium* subsp. *paratuberculosis* infected herds.
604 *Vet. Immunol. Immunopathol.* 127, 235-241.
- 605 Nagata, R., Kawaji, S., Minakawa, Y., Wang, X., Yanaka, T., Mori, Y., 2010. A specific induction of
606 interleukin-10 by the Map41 recombinant PPE antigen of *Mycobacterium avium* subsp.
607 *paratuberculosis*. *Vet. Immunol. Immunopathol.* 135, 71-78.
- 608 Nielsen, S.S., 2008. Transitions in diagnostic tests used for detection of *Mycobacterium avium* subsp.
609 *paratuberculosis* infections in cattle. *Vet. Microbiol.* 132, 274-282.
- 610 Nielsen, S.S., Ersbøll, A.K., 2006. Age at occurrence of *Mycobacterium avium* subspecies
611 *paratuberculosis* in naturally infected dairy cows. *J. Dairy Sci.* 89, 4557-4566.
- 612 Nielsen, S.S., Toft, N., 2008. Ante mortem diagnosis of paratuberculosis: A review of accuracies of
613 ELISA, interferon- γ assay and faecal culture techniques. *Vet. Microbiol.* 129, 217-237.
- 614 Nielsen, S.S., Krogh, M.A., Enevoldsen, C., 2009. Time to the occurrence of a decline in milk
615 production in cows with various paratuberculosis antibody profiles. *J. Dairy Sci.* 92, 149-155.
- 616 OIE, 2009. Standard operating procedure (SOP) for OIE validation and certification of diagnostic
617 assays. OIE, Paris, France. Available from: http://www.oie.int/vcda/eng/en_fichier_SOP.pdf
- 618 Olsen, I., Boysen, P., Kulberg, S., Hope, J.C., Jungersen, G., Storset, A.K., 2005. Bovine NK cells can
619 produce gamma interferon in response to the secreted mycobacterial proteins ESAT-6 and
620 MPP14 but not in response to MPB70. *Infect. Immun.* 73, 5628-5635.
- 621 Ransohoff, D.F., Feinstein, A.R., 1978. Problems of spectrum and bias in evaluating the efficacy of
622 diagnostic tests. *N. Eng. J. Med.* 299, 926-930.
- 623 Su, C.L., Gardner, I.A., Johnson, W.O., 2004. Diagnostic test accuracy and prevalence inferences
624 based on joint and sequential testing with finite population sampling. *Stat. Med.* 23, 2237-2255.
- 625 TDR Diagnostics Evaluation Expert Panel, 2006. Evaluation of diagnostic tests for infectious diseases:
626 general principles. *Nat. Rev. Microbiol.* 4(12 Suppl), S20-S32.
- 627 Wells, S.J., Kubat, N., Espejo, L.A., Godden, S.M., 2010. Effect of delaying exposure to Johne's
628 disease until adulthood on development of new infections in adult dairy cows. JDIP Annual
629 Conference (abstract), July 11-12, 2010, Denver, CO, USA, pp. 19.

- 630 Whiting, P., Rutjes, A.W.S., Reitsma, J.B., Bossuyt, P.M.M., Kleijnen, J., 2003. The development of
631 QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in
632 systematic reviews. *BMC Med. Res. Methodol.* 3, 25.
- 633 Whitlock, R.H., Rosenberger, A.E., Sweeney, R.W., Spencer, P.A., 1996. Distribution of *M.*
634 *paratuberculosis* in tissues of cattle from herds infected with Johne's disease. In: Chiodini RJ,
635 Hines II ME, Collins MT (eds): Proceedings of the Fifth International Colloquium on
636 Paratuberculosis, Sept. 29 - Oct. 4, 1996, Madison, Wisconsin, USA, p. 168-174.
- 637 Windsor, P.A., Whittington, R.J., 2010. Evidence for age susceptibility of cattle to Johne's disease.
638 *Vet. J.* 184, 37-44.
- 639

Table 1. Overview of study designs for diagnostic test evaluation studies for *Mycobacterium avium* subsp. *paratuberculosis* (MAP)

Study design	Description	Target conditions	Advantages	Disadvantages
Cross-sectional study with detailed pathological examination and tissue culture	Complete random sampling where enrolled animals represent the age-distribution in target population. Ideally, samples for the index test should be obtained simultaneously with collection of tissues for gross and histopathological examination, and culture or PCR.	All, but ideal for MAP infected	Pathological examination and culture/PCR of sufficient tissues is highly sensitive and specific	Costly and often not logistically feasible
Longitudinal study with follow-up using reference test	A random sample of animals should be followed for at least 4 yr, in herds with different MAP prevalences. Animals are tested frequently with a reference test such as culture of faecal samples and samples for the index test are obtained simultaneously. Animals should be followed until they are culled or to the end of study. At culling, additional tissue samples may be obtained for detection of MAP by culture or PCR.	All	Follow-up period covers typical lifetime experience in infected herds	Loss to follow-up bias; may be costly. Logistically challenging with long study periods
Cross-sectional design with complete verification: latent class analysis (LCA) of data	Data are collected using a cross-sectional design. At least two index tests with different targets (e.g. detection of agent and antibody responses) are used. Paired samples should be obtained from the same animal either simultaneously or within a short time interval. Use of LCA requires that certain assumptions be met for	All, but specifically relevant to targets of index test	Reference test need not be perfect; no assumptions necessary about its accuracy	Cross-sectional with complete verification more costly than designs with partial verification. For

inferences to be valid (Enøe et al., 2000). Partial verification approaches are possible but require a Bayesian LCA model (Su et al., 2004).

LCA, there is no explicit case definition

Hierarchical and hybrid designs	Hierarchical designs take advantage of pre-existing data about herd MAP infection status (e.g. non-infected) to define animal status. Sometimes hybrid studies are used which combine data from non-infected herds and data obtained from cross-sectional studies with complete verification in infected herds. LCA is used for inferences	Non-infected in a hierarchical design; all targets possible in a hybrid design	Efficient to estimate specificity; in hybrid designs, specificity data from non-infected herds may reduce uncertainty in inferences from LCA models	Might be difficult to find non-infected herds
Case-control	Random sampling of animals that are categorized as having the target condition of interest (cases) or not (controls) is used. The case-control study might be implemented nested within a longitudinal design. Alternately, a study might be implemented retrospectively using biologic samples that are stored in a repository.	MAP infectious; MAP affected; MAP in end-products (beef, milk etc.)	Easy; Low cost	Multiple biases (verification; spectrum) are likely especially in a retrospective design

1

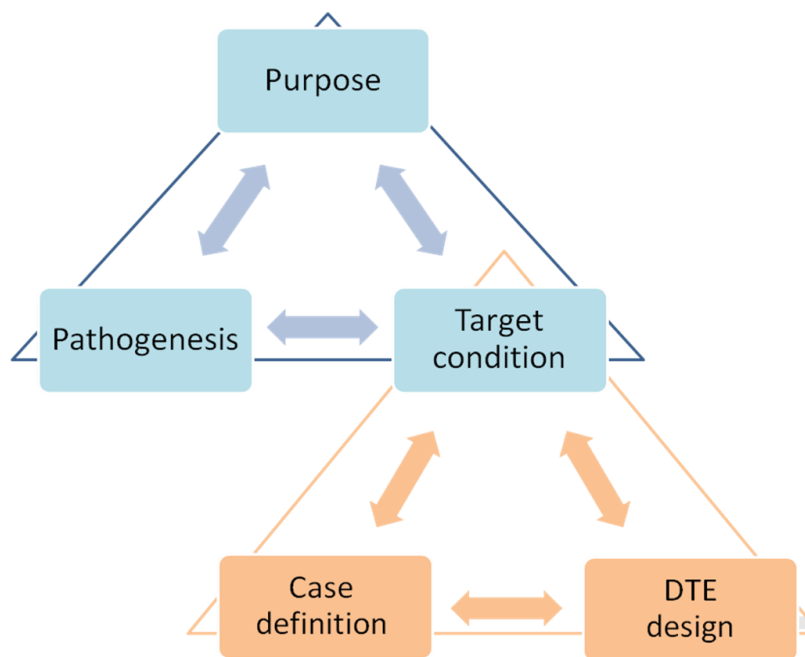


Fig. 1. Triads in a structured approach to a DTE for chronic infectious animal disease. The upper triad is the “application triad”, i.e. the context in which the test is applied should be used should be considered here, whereas the lower triad is the “evaluation triad”, where the actual DTE is created.

2

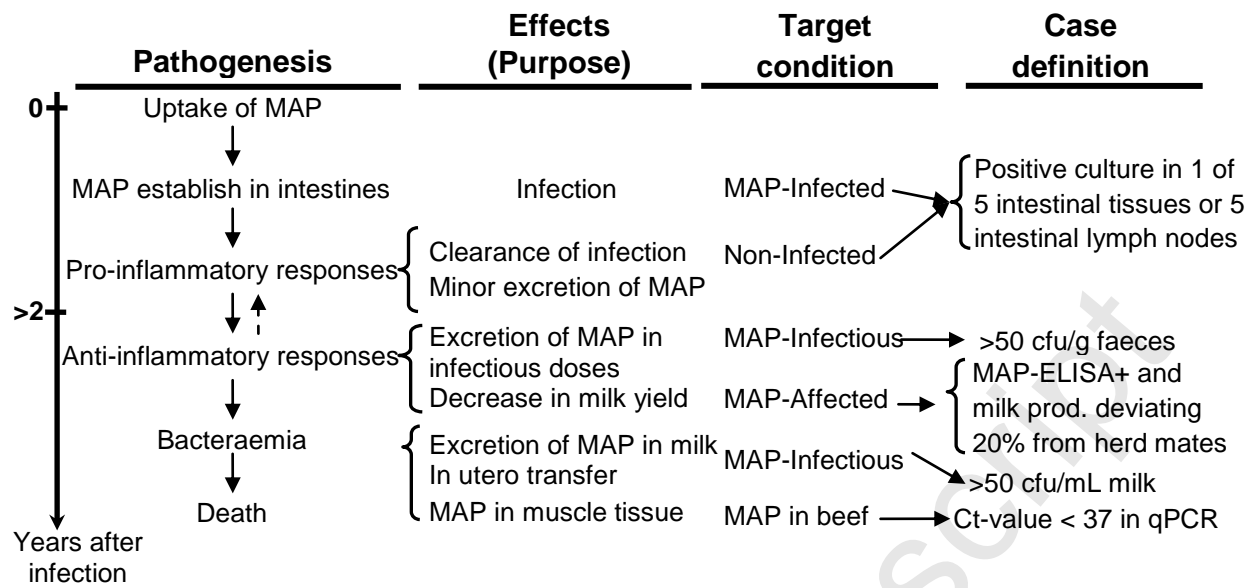


Fig. 2. Schematic presentation of hypothetical pathogenesis of *Mycobacterium avium* subsp. *paratuberculosis* (MAP) infection with resulting effects, examples of target conditions and case definitions. CFU=colony forming units; Ct = cycle threshold; qPCR = quantitative real-time PCR