

1 **Co-construction of a method for evaluating the intrinsic quality of bovine milk**  
2 **in relation to its fate**

3 L. Rey-Cadilhac<sup>a</sup>, R. Botreau<sup>a</sup>, A. Ferlay<sup>a</sup>, S. Hulin<sup>b</sup>, C. Hurtaud<sup>c</sup>, R. Lardy<sup>a</sup>, B.  
4 Martin<sup>a</sup>, C. Laurent<sup>a</sup>

5  
6 <sup>a</sup> *Université Clermont Auvergne, INRAE, VetAgro Sup, UMR Herbivores, F-63122*  
7 *Saint-Genès-Champanelle, France*

8 <sup>b</sup> *Pôle Fromager AOP Massif central, F-15000 Aurillac, France*

9 <sup>c</sup> *PEGASE, INRAE, Institut Agro, F-35590, Saint-Gilles, France*

10

11 *Corresponding author: C Laurent. E-mail: [claire.laurent@vetagro-sup.fr](mailto:claire.laurent@vetagro-sup.fr)*

12 **Abstract**

13 There are time-tested assessments for the environmental and economic aspects of  
14 sustainability. Its societal aspect has mainly been approached through the  
15 assessment of animal welfare. However, the intrinsic quality of milk is seldom taken  
16 into account. We developed a participatory construction method for the overall  
17 assessment of intrinsic milk quality in its different dimensions (sensory, technological,  
18 nutritional and health), according to the fate of the raw milk. Two assessment models  
19 were developed, for semi-skimmed standardized ultra-high temperature (UHT) milk  
20 and for pressed uncooked non-standardized raw milk cheese. They were constructed  
21 by a participatory approach involving experts in the dairy sector with the aim to obtain  
22 a diagnostic tool that could be used in the field to help farmers to manage the quality  
23 of their milk (by prioritizing improvements on major problems). They were shaped  
24 from prerequisite specifications (limited costs and time of application, desire to obtain  
25 a transparent tool with all the steps kept visible) and current technical and scientific

26 knowledge. They were based on indicators obtained from raw bulk tank milk analyses  
27 (30 for UHT milk and 50 for cheese assessments), which were then aggregated into  
28 criteria, principles, dimensions and overall intrinsic quality at farm level. The  
29 assessment models had parts in common, e.g. same four dimensions, common  
30 indicators for health and nutritional dimensions. They also had process-specific  
31 features: units chosen, criteria, indicators and weightings in relation to the final  
32 product specifications. For instance, sensory and technological dimensions are more  
33 complex and preponderant in the cheese assessment (3 principles for cheese vs. 1  
34 for UHT milk in both dimensions). Another example is the lack of microbial pathogens  
35 (as potential health risk for consumer) in the UHT milk assessment because of  
36 pasteurization. The assessment models then underwent a sensitivity analysis and an  
37 application in 30 farms in indoor and grazing periods to finally obtain overall UHT milk  
38 and cheese quality scores at a one-year level. The tool was found to be applicable at  
39 farm level. However, we observed low overall quality scores with a narrow dispersion,  
40 characteristic of a severe evaluation. Even so, the assessment models showed up  
41 seasonal differences of the UHT milk and cheese quality at both overall and  
42 dimensional levels. In the light of new scientific knowledge and future quality  
43 objectives, these are adaptable to other dairy products allowing for their specific  
44 features.

45

46 **Keywords:** Dairy cows, cheese and ultra-high temperature milk, bulk tank milk,  
47 multicriteria assessment model, participatory approach

48 **Implications**

49 This article presents the development of two intrinsic quality of milk assessment  
50 models according to milk fate (ultra-high temperature milk or raw milk cheese) and  
51 following a participatory approach. It is the first multicriteria assessment of the  
52 intrinsic quality of milk with the definition of principles to evaluate four quality  
53 dimensions (sensory, technological, nutritional, and health), criteria and indicators to  
54 be considered, and rules to interpret them. The approach followed is adaptable to  
55 other dairy products, allowing for the specific features of the product and the  
56 demands of the stakeholder group seeking to develop it.

## 57 **Introduction**

58 In a context of a continuous decrease in the number of dairy farms, their assessment  
59 for sustainability is a crucial issue. There are time-tested assessment methods for the  
60 environmental and economic aspects of sustainability (Halberg et al., 2005; Zahm et  
61 al., 2008), but the assessment of its social aspect is more complex. Social  
62 sustainability has two levels: social sustainability within the farm (quality of life,  
63 working conditions...) and social sustainability external to the farm, constantly  
64 evolving with values and concerns of the society (Lebacqz et al., 2013): the main  
65 societal expectations are animal welfare and intrinsic product quality. There has been  
66 a significant development of tools for assessing animal welfare (Botreau et al., 2008),  
67 but the intrinsic quality of milk is seldom taken into account in assessment methods  
68 (Lebacqz et al., 2013). A few evaluation methods for the intrinsic quality of milk have  
69 been proposed (Harris, 1998; Müller-Lindenlauf et al., 2010; Zucali et al., 2016) with  
70 few chosen indicators (3–8 depending on the method), and mostly related to health  
71 (somatic cells count, protein and fat contents, milk fat composition or total bacterial  
72 count), and technological dimensions (microorganisms) of quality. Intrinsic quality can

73 be defined as a combination of sensory, technological, nutritional, and hygiene  
74 dimensions (Coulon, 2008) and the quality expectations are not the same according  
75 to the fate of the milk (Prache et al., 2020). However, these existing evaluations  
76 address quality on targeted aspects and without considering the fate of the milk. The  
77 aim of this study was to develop a participatory construction method for the overall  
78 assessment of intrinsic milk quality taking into account all its dimensions depending  
79 on the fate of the raw milk. Ultimately, the goal is to obtain an assessment model that  
80 could be used in the field to help farmers to manage the quality of their milk.

## 81 **Material and methods**

82 The multicriteria assessment model of the intrinsic quality of milk was constructed in  
83 three steps (Lairez et al., 2015) from the participative approach : definition of the  
84 scope, definition of the conceptual and methodological frameworks and validation of  
85 the assessment model (Figure 1).

### 86 ***The different steps of the assessment model construction***

#### 87 *Step 1 - Definition of the scope of the assessment*

88 The first step was the definition of the scope of the assessment by the experts,  
89 described by the object to be evaluated and its spatial and temporal scale, the aim of  
90 the assessment and its future users, and its prerequisites (e.g. financial and time  
91 costs, ability to run it).

#### 92 *Step 2a - Definition of the conceptual framework*

93 The overall quality had to be defined in several *dimensions* or qualities, then  
94 subdivided into *principles* and *criteria* (and *subcriteria* if needed) that itemized the  
95 principles into more concrete categories. Criteria were assessed by *indicators* that

96 were measurable (= raw data) and that had to meet four conditions (Lairez et al.,  
97 2015): scientific relevance, feasibility, clarity and interpretability. We then checked  
98 that the hierarchical structure respected the scope choices (step 1; Figure 1),  
99 especially the indicators that had to fulfil strict specifications in terms of feasibility and  
100 cost. If possible or necessary we then simplified the structure in that sense.

#### 101 *Step 2b - Definition of the methodological framework*

102 The methodological framework of the assessment is defined from the previously built  
103 hierarchical structure to interpret and aggregate the indicators up to the overall  
104 assessment (Figure 1). We decided to use the CONTRA tool, an aggregation method  
105 based on a decision tree using fuzzy sets (Bockstaller et al., 2017; Botreau et al.,  
106 2018). This tool requires the setting of a few parameters to interpret and to aggregate  
107 the indicators up to the overall assessment. In our case, a common quantitative scale  
108 for the scoring was defined: from 0, worst possible quality score, to 10, best possible  
109 quality score.

110 Interpretation of the indicators. For each indicator, threshold values to delimit the  
111 favourable class (above which the score will be 10/10) and the unfavourable one  
112 (below which the score will be 0/10) had to be defined. In the fuzzy zone (between  
113 the two classes), the element to be interpreted (indicator value, criterion or principle  
114 score, generically called 'input variable') is considered to be part of both classes with  
115 different degrees of membership. CONTRA then aggregates these two membership  
116 degrees to produce an intermediate score for each input (Bockstaller et al., 2017),  
117 thus limiting the risk of a threshold effect induced by the drop from one class to  
118 another.

119 Aggregation of scores. CONTRA aggregates indicators' scores to produce the  
120 criterion score (and so on until the overall quality score). This aggregation relies both

121 on weightings (is one indicator more important than another?) and management of  
122 compensations between indicators (do we want to be severe by limiting the possibility  
123 of compensating for a bad score on one indicator by a good one on another?). Other  
124 specific settings that are not presented here were also defined (refer to Bockstaller et  
125 al., 2017).

126 Thus, three types of parameter had to be defined: threshold values, weights and  
127 rules for compensation management.

### 128 *Step 3 - Validation of the assessment model*

129 Sensitivity analysis. For a better overview of the assessment tool's behaviour, a  
130 sensitivity analysis was performed using the screening method initially developed by  
131 Morris (1991) and modified by Campolongo et al., (2007), known as the elementary  
132 effects method. For each input variable, the method calculates two indices,  $\mu$   
133 (assessing the overall influence of the input on the final result of the assessment) and  
134  $\sigma$  (estimating both the nonlinearity of the impact and interactions with other inputs). A  
135 high value of  $\mu$  indicates a marked overall influence on the output (total effect). A high  
136  $\sigma$  indicates either a non-linear effect on the output, or involvement in interaction with  
137 other inputs (higher than one-order effects). More details are given in Supplementary  
138 Material S1.

139 Test of the assessment model on dairy farms. In addition, the behaviour of the  
140 models was also checked through a test on 30 commercial farms. Farms  
141 characteristics are presented in Table S1, they were selected to cover a large  
142 diversity of farming systems (plain and mountain, intensive and extensive...). Bulk  
143 tank milk samples were collected in July (grazing period) and in January-March  
144 (indoor period) in a tank containing an even number of milkings (2, 4 or 6) except in  
145 two farms equipped with a milking robot. A 1.2 L sample of milk was directly collected

146 in the tank after 5 min agitation using a sterile sampling rod through the manhole.  
147 The sample was kept between 0 and 4°C during the transfer to the lab within 18 h.  
148 Twelve and nine indicators were analysed on fresh milk within 48 h after sampling  
149 and on frozen milk samples (–20°C), respectively (Table S2). Some of the analyses  
150 were performed by private laboratories, using mostly routine methods. The others  
151 were performed by research laboratories using reference methods (e.g. gas  
152 chromatography for milk fatty acids, FA). When data were missing (some analyses  
153 were not carried out because sampling and analysis of bulk tank milks occurred  
154 before the assessment model had been finalized), they were approximated by  
155 existing equations or assigned a unique mean value for all the farms obtained from  
156 the literature or databases (Table S2). The two assessment models were then  
157 applied to those 30 farms.

### 158 ***The participative approach for the construction of the multicriteria assessment*** 159 ***model***

160 The construction of the assessment model was based on a participative approach  
161 with the consultation of scientists and stakeholders of the dairy sector. This approach  
162 is based on the confrontation of diverse points of view to reach a consensus that  
163 satisfies the whole group (bottom-up approach) within focus group meetings, and so  
164 allows a better adoption of the tool by the stakeholders who helped to build it. The  
165 people involved in the process, later called ‘experts’, were divided into 2 groups: (i)  
166 Leader group composed of researchers on milk quality (2 people) and stakeholders  
167 (2): representative of the cheese sector and one from a private company. They  
168 participated in the specific steps identified in Figure 1; (ii) Specific experts group  
169 according to the indicators addressed, who participated in one or more ad hoc

170 meetings composed of researchers on specific milk compounds (2) or on milk  
171 technology (2), technologists (1), doctors and nutritionists (4), farm advisors (2),  
172 representatives of the cheese sector (2), process manager within a dairy company  
173 (1). They participated in the specific steps identified in Figure 1.  
174 The focus group meetings (about 4 h long) were organized following the simplified  
175 DELPHI (Linstone and Turoff, 1975) method with only one round with phases of  
176 individual reflection of experts and phases of pooling/discussion (for more details,  
177 see Supplementary material S2).

## 178 **Results**

### 179 ***Definition of the scope of the assessment***

180 During the first meeting, the experts defined the object to be evaluated as the bulk  
181 tank milk of a farm, safe for human consumption and thus complying with the relevant  
182 legislation). The intrinsic quality of bulk tank milk from the farm was evaluated at the  
183 year level.

184 The aim of the assessment was to propose an action-oriented diagnostic tool to be  
185 used by farmers to manage the overall quality of their milk (by prioritizing  
186 improvements on major problems). This had an impact on the setting of the following  
187 initial prerequisites by the experts: (i) transparency: to keep all the intermediary  
188 steps accessible and to make it possible to point out the strengths and weaknesses  
189 of the assessment, (ii) genericity and standardization: the assessment should be  
190 applicable to any dairy farm, independently of the local context, (iii) use in routine  
191 practice: this implies a limited cost (<100 €/year), a limited time for  
192 sampling/surveying (<1 h), the use of routine methods for milk analyses (chemical,  
193 microbial or spectral methods), and a time limit for processing the information (<1 h).



194 The sampling must thus be easy (a single milk sample taken from the bulk) and non-  
195 invasive (no measure on animals, only on milk).

196 To obtain an assessment at year level but to limit the costs, the experts decided to  
197 plan only two periods of assessment: grazing and indoor. The overall score at year  
198 level would be calculated from a weighted average of the two seasonal quality scores  
199 as follows:

200 Year quality score = [Annual part time spent in barn (%) × Indoor period quality  
201 score + Annual part time spent on pasture (%) × Grazing period quality score] / 100

202 The experts soon found that it was not possible to describe the dimensions at lower  
203 levels and define criteria of quality without taking into account the fate of the milk.

204 Accordingly, two different products were targeted, and two assessment models were  
205 developed: one for semi-skimmed standardized ultra-high temperature (UHT) milk  
206 and the other for pressed uncooked non-standardized raw milk cheese, these having  
207 very different transformation processes and final desired characteristics.

## 208 ***Definition of the conceptual framework***

209 *Intrinsic quality defined through its different dimensions and according to the fate of*  
210 *the milk*

211 The experts defined intrinsic milk quality as a combination of four major dimensions  
212 during the first meeting too: (i) sensory: ability of the milk to contribute to the  
213 organoleptic characteristics of the final product, (ii) technological: ability of the milk to  
214 be correctly processed into the final product, (iii) nutritional: ability of the milk to give  
215 a final product that contributes significantly to the coverage of consumers' nutritional  
216 needs, and (iv) health: ability of the milk to give a final product that has potentially  
217 more beneficial and less deleterious effects on human health. The contribution of milk

218 (bottled or carton, flavoured, concentrated or powdered milk and other milks) and  
219 cheese (ripened, processed, unripened fresh or other cheeses) to the total  
220 recommended daily intakes was calculated from data on recommended nutrient  
221 intakes by nutrient, the average intake of milk and cheese and the levels of these  
222 nutrients in milk and cheese (Agence française de sécurité sanitaire des aliments  
223 (AFSSA), 2009). Only those nutrients for which the contribution of milk and cheese  
224 exceeded 5% of the total daily intakes were kept in the nutritional quality part of the  
225 respective assessments. Constituents whose concentrations can be greatly  
226 increased or decreased by the process were also excluded. For instance, vitamins B  
227 were not taken into account in the cheese assessment even though cheese supplies  
228 them in abundance, these vitamins resulting mainly from synthesis by cheese flora  
229 (Reif et al., 1976) and not from bulk milk content.

### 230 *From dimensions to indicators*

231 Twenty-one meetings were needed to define the principles, criteria and indicators  
232 (together with their interpretation) for the two assessment models.  
233 This construction was an iterative process, particularly in relation to the choice of  
234 indicators, which had to be both scientifically relevant and compliant with the  
235 specifications (limited cost and easy analysis, in particular). The hierarchical structure  
236 was thus simplified and the Leader group decided to do it with 3 possibilities for the  
237 indicators originally selected: (i) if the indicator was measurable (analytical cost and  
238 method available in private laboratories), then it was kept; (ii) if the indicator was not  
239 directly easily measurable but it could be predicted by analytical methods (e.g.  
240 prediction by mid-infrared method) or proxies that allowed to approach it (e.g.  
241 coagulase positive staphylococcus level to estimate biological toxins level  
242 (enterotoxins), *pseudomonas* level to estimate coming proteolysis intensity), then it

243 was kept too; (iii) if the indicator was not directly measurable and there was no  
244 simplified analysis method or proxy that allowed to approach it, then it was deleted. If  
245 no indicator was retained for a branch of the structure, then the entire branch was  
246 deleted.

247 The UHT milk conceptual framework was structured into 4 dimensions, 9 principles  
248 (1–4 per dimension), 27 criteria (plus 13 sublevels), and 28 indicators (Figure 2).

249 The cheese conceptual framework was structured into 4 dimensions, 11 principles  
250 (2–3 per dimension), 27 criteria (plus 15 sublevels and 6 sub-sublevels), and 50  
251 indicators (Figure 3).

## 252 ***Definition of the methodological framework***

### 253 *Interpretation of indicators*

254 For a majority of indicators, the thresholds were defined during the consultation  
255 process. However, for most of the nutritional and health indicators no consensus was  
256 reached and experts were only able to define the global sense of interpretation (e.g.  
257 lower is better) but not the thresholds. These were therefore defined from the  
258 analysis of a database of around 1300 individuals (Coppa et al., 2013; Chassaing et  
259 al., 2016): the 0/10 score was defined as the 2.5th percentile and the 10/10 score as  
260 the 97.5th percentile of the values of this database for indicators interpreted as  
261 'higher is better' (and the opposite for 'lower is better' indicators), and then discussed  
262 and validated with the experts (Table S3).

### 263 *Aggregation of inputs*

264 Nine meetings were then needed to aggregate indicators, criteria, principles and  
265 dimensions: definition of weightings (7 meetings) and compensation rules (2  
266 meetings) by the experts.

267 Weightings. The UHT milk assessment model was principally based on nutritional  
268 quality (weighting of 40%), which took into account the mineral, protein, and vitamin  
269 inputs. Then came health quality (30%), predominantly based on the impact of lipid  
270 content on health. Lastly, the technological (20%) and sensory (10%) qualities were  
271 characterized by only one principle each: stability and suitability for preservation, and  
272 taste, respectively. The cheese assessment model was largely based on sensory  
273 (35%) and technological (30%) qualities. Sensory quality was mainly based on  
274 flavour (flavour defects from non-desirable microorganisms) of the final cheese.  
275 Technological quality was predominantly evaluated by the cheese yield capacity and  
276 drainability. Health quality (20%) was mainly based on the effects of the bioactive  
277 components (especially lipids) and nutritional quality (15%) based on energy and  
278 macro-element inputs. All weightings for UHT milk and cheese are given in Figures 2  
279 and 3, respectively.

280 Compensations. The decision was collectively made to limit compensations to  
281 promote an overall quality without bad results on some inputs. Two main rules were  
282 defined by the experts whatever the final product:  
283 (i) the more heterogeneous the input scores, the more limited the compensation. This  
284 means that with the same average score, the more heterogeneous the milk was over  
285 the inputs to be aggregated, the lower was the aggregate score;  
286 (ii) the lower the input scores, the more limited the compensation: higher severity  
287 when globally input scores are bad.

## 288 ***Validation of the assessments***

### 289 *Sensitivity analysis of the assessment models*

290 The results obtained on the seven most influential indicators on the overall score for  
291 both models are presented in Table 1 (for more details, see Tables S4 and S5). For  
292 both assessment models, the  $\mu$  were all of the same order of magnitude so no  
293 indicators predominated in their influence on the overall score. However, some  
294 indicators were more influential than others. The most influential indicator in both  
295 assessments was the *Pseudomonas* level. This indicator was present and influential  
296 in all four dimensions of the two assessments. It was the only indicator for sensory  
297 dimension for milk assessment and the most influential in the sensory dimension of  
298 the cheese assessment.

299 The other most influential indicators of the overall UHT milk quality score were initial  
300 pH, calcium,  $\beta$ -carotenes, vitamin B2 and lactose concentrations. They were also  
301 among the most influential indicators of the main dimensions: calcium and vitamin B2  
302 concentrations for nutritional dimension score;  $\beta$ -carotenes and lactose  
303 concentrations for health dimension score, and initial pH and *Pseudomonas* level for  
304 technological dimension score.

305 The other most influential indicators of the overall cheese quality score were the  
306 casein concentration, total bacteria count,  $\alpha$ -linolenic acid (ALA), urea and calcium  
307 concentrations. Caseins were present in three of the four dimensions, health,  
308 nutritional and technological, and the most influential indicator of both these last  
309 dimension scores. The other most influential factor in the cheese nutritional  
310 dimension score was the ALA concentration, whereas it was urea concentration for  
311 the cheese technological dimension score. For the sensory dimension score,  
312 *Pseudomonas* level and the total bacteria count were the most influential indicators.  
313 The most influential indicators of the health dimension score were not included in the  
314 most influential indicators of the cheese overall score but were  $\beta$ -glucuronidase-

315 positive E. coli level ( $\mu = 1.78$ ,  $\sigma = 4.00$ ; Table S5) and  $\beta$ -carotene concentration  
316 ( $\mu = 1.7$ ,  $\sigma = 0.86$ ; Table S5).

### 317 *Test of the assessment models on dairy farms*

318 The ranges of values obtained for indicators from the milks collected in the 30 farms  
319 are presented in Table S3, covering fairly well the range defined by the thresholds for  
320 each indicator. Mean quality scores of the 30 farms are presented in Table 2 and in  
321 Figures S1 and S2.

322 At year level, cheese and UHT milk global quality scores were of  $3.5 \pm 0.7$  and  $2.6 \pm$   
323  $0.5$  out of 10 (mean  $\pm$  SD), respectively, with a narrow variability and a maximum that  
324 did not exceed 5 for both assessment models. At dimension and season scales, the  
325 variability increased, with a broader observed distribution of the scores. When  
326 considered at season level, quality scores were higher in grazing than indoors for the  
327 cheese and UHT milk assessments ( $3.7 \pm 1.0$  vs.  $3.1 \pm 1.0$  for cheese,  $p = 0.02$ ;  $2.8$   
328  $\pm 0.5$  vs.  $2.1 \pm 0.4$  for UHT milk,  $p < 0.001$ ). At dimensional level, health scores were  
329 significantly higher in the grazing period, whereas no difference by season in  
330 nutritional scores appeared in either assessment. Also, technological scores were  
331 higher in the grazing period in the UHT milk assessment model but no differences  
332 across seasons appeared for the cheese assessment model. Finally, sensory scores  
333 were higher in the grazing period in the cheese assessment (sensory score in the  
334 UHT milk assessment was assigned the same value for all the farms because of  
335 missing data, so no comparison could be made).

## 336 **Discussion**

### 337 ***Definition of intrinsic quality***

338 Milk intrinsic quality has often been defined as the resultant of four dimensions:  
339 nutritional, sanitary/safety/hygiene, sensory, and technological (Coulon, 2008; García  
340 et al., 2014). In our assessment models, the sensory and technological dimensions  
341 have a meaning close to that usually taken in the literature, but the nutritional and  
342 health dimensions encompass a different meaning. Nutritional quality usually refers  
343 to the combination of ability to meet food requirements and positive/deleterious  
344 effects of milk on human health. For example, FA profile has often been related to the  
345 nutritional quality of milk, although the human nutritional needs for FAs can be strictly  
346 defined only for essential FAs: ALA and linoleic acid (LA), and not on the other FAs  
347 having beneficial or deleterious effect on health (Agence nationale de sécurité  
348 sanitaire de l'alimentation, de l'environnement et du travail (ANSES), 2011). In this  
349 light, the experts decided to clearly separate nutritional aspects from effects on health  
350 in these assessment models. Only the ALA and the LA contents were therefore taken  
351 into account in the nutritional dimension (in the cheese assessment model). Finally,  
352 sanitary quality usually refers to the absence of hazardousness often evaluated as  
353 the absence of pathogens and/or of chemical pollutants in the milk (Coulon, 2008;  
354 García et al., 2014). Here, the definition was enlarged to cover all potential impacts  
355 on human health, positive and negative. As the appraised object was a bulk tank milk  
356 safe for human consumption, pathogens such as *Listeria monocytogenes* did not  
357 appear in our assessment models despite belonging to health aspects.

### 358 ***Assessment models according to the fate of milk***

359 Considering the fate of the milk entailed differences between the two assessment  
360 models of the quality of the milk (for cheese and UHT milk).

### 361 *Importance of dimensions according to the fate of milk*

362 The experts opted to assign more weight to the sensory aspects for cheese  
363 assessment model, considering that the raw milk cheese is purchased by consumers  
364 for its gustative and olfactory experience and less for nutritional or health  
365 expectations. As the technological process is very complex and largely responsible,  
366 in interaction with the milk composition (Fréтин et al., 2019), for the sensory  
367 characteristics of the final cheese, the technological dimension was also given a  
368 large weighting. On the contrary, UHT milk is an 'everyday' product that is mainly  
369 used as an ingredient among others, so that the sensory dimension in UHT milk  
370 assessment model is limited to its taste or rather the absence of defects in the taste,  
371 as milk is defined as having no special taste per se (Clark and Bodyfelt, 2009). As it  
372 is more frequently consumed, the expectations of its effects on nutrition and health  
373 were assumed to be high, justifying higher assigned weighting.

#### 374 *Criteria and indicators according to the fate of milk*

375 At criteria and indicator levels, the differences between the two assessment models  
376 were still perceptive. The sensory dimension was based only on the absence of  
377 defects in the UHT milk assessment, while for the cheese, three aspects referring to  
378 the human senses were considered: appearance, texture, and flavour. The  
379 technological dimension of the UHT milk assessment model was based on the  
380 capacity of the milk to permit good conservation of the final product, linked to acidity  
381 and micellar stability (linked in turn to possible milk precipitation during storage;  
382 Gaucher et al., 2011). In comparison, the technological dimension in the cheese  
383 assessment model referred more to parameters that could affect the transformation  
384 process, such as acidification capacity, drainability and more generally cheese yield  
385 capacity.



386 The transformation process leading to the final product also meant differences  
387 between indicators. For example, no microbial aspects were considered in the UHT  
388 milk assessment model as the microbial constituents are destroyed by the process  
389 (heating at ultra-high temperature), except for the level of *Pseudomonas*, considered  
390 as a predictor of proteolysis that can occur both before and after the transformation  
391 process. This group of bacteria produces proteolytic enzymes able to stay active after  
392 a high temperature treatment and impair milk quality (Gaucher et al., 2011). Similarly,  
393 lactose and soluble protein contents were not relevant to the cheese assessment  
394 model as they are nearly completely released in the whey during the clotting and  
395 draining stages of the pressed uncooked non-standardized raw milk cheese process  
396 (Kindstedt, 2014). On the contrary, FA profile, minerals, and total  $\beta$ -carotenes  
397 appeared in both assessment models.

398 *Indicator units and thresholds according to the fate of milk and the dimension*

399 FA contents were measured as concentrations (g/L of milk) in the cheese  
400 assessment model, as the milk used in the process was non-standardized for milk  
401 fat. Conversely, in the UHT milk process, the milk was fat-standardized, milks were  
402 not differentiated by FA concentrations, but instead by FA relative proportions (% of  
403 total FAs).

404 For a given indicator used in different dimensions and in both assessment models,  
405 thresholds can differ from one dimension to another and between models. For  
406 example, in the cheese assessment model, fat-to-protein ratio was used as an  
407 indicator for the cheese yield in the technological dimension and for the cheese  
408 texture in the sensory dimension, with different interpretations according to the  
409 dimension. The main drivers of the cheese yield are proteins (more specifically  
410 caseins) and fat content in milk, with a higher weighting for the protein part (Banks,

411 2007). However, a certain balance between fat and protein is required to guarantee  
412 the ability of the protein matrix to retain fat globules during cutting and early stages of  
413 stirring and limit fat loss in whey (Guinee et al., 2007). The ‘technological’ score of  
414 the fat-to-protein ratio was therefore higher for a lower ratio (10/10 for ratio  $\leq 1.10$  and  
415 0/10 for a ratio  $\geq 1.40$ ; Table S3). The process is different for the effect of the fat-to-  
416 protein ratio on the cheese texture. Cheese firmness is closely related to fat in dry  
417 matter, which is linked to the fat-to-protein ratio of the milk. An increase in the fat-to-  
418 protein ratio will thus give a softer, smoother cheese (Coulon et al., 2004).  
419 Cheesemaking targets an intermediate fat-to-protein ratio, so that the cheese texture  
420 corresponds to the standards for the cheese variety. Hence the distribution of  
421 ‘sensory’ scores was bell-shaped, with an optimal score for a ratio of 1.15, and then  
422 degressive scores for larger or smaller ratios (0/10 for ratio  $\leq 1.05$  and  $\geq 1.35$ ).  
423 To conclude, of the 37 and the 21 different indicators in the cheese and the milk  
424 assessment models, respectively, 7 indicators are common (same unit, same  
425 interpretation) between the two models.

#### 426 ***Assessment models shaped from participative approach***

427 The building process, using a participative approach, produced the first overall  
428 intrinsic milk quality assessment model, to our knowledge. It made it possible to  
429 reconcile the points of view of the different actors involved in the development of the  
430 models. Scientists from the dairy field and professionals from the dairy industry,  
431 ensured the scientific and technical validity of the information and thus the credibility  
432 of the assessment models. Perhaps the actors missing from this panel who could  
433 confer a stronger legitimacy are the breeders. Besides, the number of meetings  
434 required was high but could perhaps be reduced with experience. However, reducing

435 this number could negatively impact the participatory approach and reduce credibility  
436 of the tool (e.g. through lack of consensus). The models were salient since the  
437 information given by the assessment was relevant for the decision makers (farm  
438 level) and appropriate to the context (suitable for use in the field and adapted to two  
439 products). According to Cash et al. (2002), salience, credibility and legitimacy are the  
440 three properties that an assessment must possess to be functional as a decision  
441 support tool and be accepted by the stakeholders.

#### 442 *Assessment models shaped from the scope*

443 Experts chose the indicators associated with each dimension and prioritized them  
444 considering the constraints of the specifications identified for the tool at the beginning  
445 of the project. Some relevant indicators were withdrawn because they were too costly  
446 to analyse. Some other indicators were not directly measured, but estimated, either  
447 by others with less costly analyses, or by mid-infrared spectrometry or flow cytometry  
448 methods when reliable prediction equations existed (Table S2). The health dimension  
449 was particularly impacted by these specifications by ignoring some pathogens, trace  
450 metals and organic pollutants, for which there are threshold values not to be  
451 exceeded to allow the commercialization of the product.

452 Some other indicators were kept in the assessment models even if they could only be  
453 analysed with reference analysis methods (so they could not be analysed on all  
454 laboratories and had a quite high cost) because experts considered that in the near  
455 future routine analysis methods will be developed. This induced a gap between the  
456 initial specifications and the developed assessment models on these points (that  
457 affected the test on 30 farms). This will have to be updated in the future according to  
458 scientific and technical advances in analytical methods.

459 Beyond of the choice of the relevant indicators, the scope of the assessment also  
460 impacted the developed models in their structure in order to maintain the information  
461 available and transparent at all levels, from indicators to the overall assessment. This  
462 allows the models to be used to advise the farmers (by self-assessment or with the  
463 help of farm advisor) by identifying the main advantages (to be maintained) and  
464 drawbacks (to be improved) in terms of milk quality. An example of such an  
465 application on a real farm is given in Figure S3. However, this is only a diagnostic  
466 tool, to find the best remedial solutions the farmer or his advisor would have to think  
467 about new practices that can solve drawbacks without negatively impacting the good  
468 points observed in terms of quality.

469 *Assessment models shaped from current technical and scientific knowledge and its*  
470 *interpretation by the focus groups*

471 Experts chose the indicators with their current knowledge on their relevance and the  
472 possibility of defining thresholds to evaluate them (according to the knowledge and  
473 the references available at a given time). As knowledge and references evolve  
474 continuously, the assessments had to be upgradeable. For the health and nutritional  
475 dimensions of quality, the panel of experts agreed on the choice of the indicators and  
476 on their rough interpretation (such as, for example, the higher the ALA content, the  
477 better the health evaluation of the milk, in accordance with the literature (Haug et al.,  
478 2007; ANSES, 2011), but they did not succeed in giving threshold values for most of  
479 the indicators, although they successfully defined them for sensory and technological  
480 dimensions. Hence it was more difficult to determine a threshold value above which  
481 the quality was impaired in the health and nutritional dimensions because of a lack of  
482 consensus among the experts due to insufficient knowledge. Levels of nutrients from  
483 milk that could affect health and nutrition positively or negatively are very difficult to

484 determine as their interpretation depends on the composition of the rest of the  
485 consumer's diet and on many other aspects (physical activity, genetics, environment,  
486 etc.). Quantitative recommendations are set only on the overall consumption of the  
487 constituents (minerals, FAs, etc.) (Agence nationale de sécurité sanitaire de  
488 l'alimentation, de l'environnement et du travail (ANSES), 2011), and not on the  
489 consumption of dairy products and even less on the nutrient content in the dairy  
490 product itself (Martin et al., 2019). These dimensions of milk quality are also more  
491 sensitive issues, for which the positioning of the experts was more difficult. A  
492 database was therefore used to define these thresholds values. This means that they  
493 do not have a direct nutritional or health meaning (no thresholds for which the  
494 indicator would become dangerous), even if the general orientation of the  
495 interpretation is known. It will be possible to review this point in step with the future  
496 evolution of scientific knowledge on the contribution of dairy products to dietary  
497 recommendations and requirements (Martin et al., 2019). In addition, the indicators  
498 used to assess nutritional quality were selected among those providing at least 5% of  
499 needs by the products concerned (UHT milk and cheese) thus depending on the  
500 current, geographically targeted (France) consumption of these products. As this  
501 consumption can vary according to place and time, it had to be possible to modulate  
502 and update the assessment.

503 The participatory approach also allowed stimulating discussion among the experts in  
504 the collective focus group times and produced unexpected ways to measure some  
505 milk quality criteria. For example, the protein level was evaluated by both protein  
506 content and the level of *Pseudomonas*, which can induce proteolysis before milk  
507 transformation that would lead to denaturation of the proteins, impairing coagulation.  
508 The alcohol and Ramsdell tests are rarely used as indicators in the literature, but are

509 commonly used in the dairy industry to test the stability of the colloidal suspension of  
510 caseins and thus the stability of milk (Gaucher, 2007).

### 511 ***Validation of the assessments***

#### 512 *Pseudomonas, an influent indicator*

513 In both assessment models, the level of Pseudomonas appear to be most influential  
514 input (Table 1). As previously explained this indicator is considered as a proxy of  
515 proteolysis that will occur between the measurement in the tank and the milk delivery  
516 to the transformation plant. Whenever an indicator related to the proteins content in  
517 milk (total proteins, caseins, soluble proteins) was used, the experts associated it  
518 with the Pseudomonas indicator to consider a potential level of proteolysis. As a  
519 consequence, this indicator is used several times for both assessments. Its great  
520 influence on the results is thus not surprising but may be questioned: protein  
521 degradability is obviously relevant in respect with the milk quality, but does it deserve  
522 to have such an important overall weight? Deeper investigations should be performed  
523 to answer that question and it would be possible to revise the models accordingly.

#### 524 *A low but real discriminating power*

525 The assessment was performed on 30 farms. Overall quality scores on the 30 farms  
526 presented a narrow dispersion, which could question the power of the assessment to  
527 discriminate the bulk tank milks of the different farms in terms of intrinsic quality.  
528 However, some indicators could not be measured (alcohol test, Dornic acidity, levels  
529 of Pseudomonas and E. coli) and the corresponding missing values were replaced by  
530 the same mean value extracted from databases or literature (Table S2) for all the  
531 farms, independently of the season, which could explain this narrow dispersion.  
532 Despite this limitation, the assessment was able to discriminate the different milks in

533 terms of intrinsic quality at the seasonal level as it showed significant differences  
534 between indoor and grazing period quality scores of the cheese and UHT milk  
535 assessments. Quality scores were better in grazing than in indoor periods, resulting  
536 from significant seasonal differences in health and sensory scores, mainly in the  
537 cheese assessment, and in health and technological scores in the UHT milk  
538 assessment. This supports the choice of the experts to plan two periods of  
539 assessment to take this seasonal variability into account in the final score. Milk  
540 components and properties related to nutritional, technological or organoleptic  
541 aspects are known to vary widely during the year, notably from changes in the  
542 feeding system (Verdier-Metz et al., 2005; Ferlay et al., 2006; Hurtaud et al., 2014).

#### 543 *A severe assessment tool*

544 The choice to limit compensation between scores led to low overall scores. To obtain  
545 a good score at each aggregation level, the milk of a farm must have good scores in  
546 all sublevel parts. This severity was accentuated by the construction of the  
547 assessments. Some indicators such as *Pseudomonas* level or casein concentrations  
548 served several times in the assessment to evaluate different criteria and principles,  
549 resulting in a high influence of these indicators on the assessments (Table 1). Thus, if  
550 a bulk tank milk has a bad score on this indicator, it will be penalised several times.

#### 551 **Conclusion and perspectives**

552 An innovative method for developing milk quality assessment has been built and led  
553 to an evaluation of the intrinsic quality of milk taking into account its different  
554 dimensions (nutritional, health, technological and sensory dimensions) as well as the  
555 fate of the milk (semi-skimmed standardized UHT milk or pressed uncooked non-  
556 standardized raw milk cheese). These two assessment models have the same four

557 dimensions but differ greatly on principles (mainly on sensory and technological  
558 ones), criteria and indicators as well as on weights given to the inputs. They were  
559 based on the decisions of experts, according to their own knowledge and opinions.  
560 The models built are intended to be evolutive and the different elements can be  
561 revised: addition or deletion of indicators, definition of thresholds, weights assigned  
562 to inputs, compensation rules, in order to provide improvement at each level of the  
563 evaluation process.

564 Moreover, this method can be adapted to other products (particularly dairy products)  
565 according to their specificities and constraints in use (costs, application time and  
566 users). The assessment models are transparent. They allow the assessor to start  
567 from the overall score and access all the intermediate scores up to the indicators  
568 measured in the milk. It is therefore possible to identify, at each level, what can be  
569 maintained and what has to be improved. Combined with a thorough knowledge of  
570 the relationships among the inputs and the breeding practices that influence them,  
571 these models can be used to instigate changes in farming practices to improve milk  
572 quality.

### 573 **Ethics approval**

574 No animal were used.

### 575 **Data and model availability statement**

576 None of the data were deposited in an official repository. The rights of access to the  
577 data, or models will be made available by the respective author upon a confidential  
578 request.

### 579 **Author ORCIDs**



580 L. Rey-Cadilhac : 0000-0002-4554-7662

581 R. Botreau : 0000-0002-9599-4313

582 A. Ferlay : 0000-0002-0651-792X

583 S. Hulin :

584 C. Hurtaud : 0000-0002-3965-826X

585 R. Lardy : 0000-0003-1338-8553

586 B. Martin : 0000-0003-2501-8306

587 C. Laurent : 0000-0002-8350-8496

588 **Author contributions**

589 L. Rey-Cadilhac : Validation, Formal analysis, Writing - Original Draft, Writing -

590 Review & Editing, Visualization

591 R. Botreau : Conceptualization, Methodology, Software, Formal analysis,

592 Investigation, Writing - Original Draft, Supervision, Project administration, Funding

593 acquisition

594 A. Ferlay : Resources, Writing - Original Draft, Supervision

595 S. Hulin : Conceptualization, Writing - Review & Editing, Project administration

596 C. Hurtaud : Resources, Writing - Review & Editing

597 R. Lardy : Validation, Formal analysis, Writing - Review & Editing

598 B. Martin : Conceptualization, Methodology, Writing - Review & Editing

599 C. Laurent : Conceptualization, Methodology, Data Curation, Investigation, Writing -

600 Original Draft, Supervision, Project administration, Funding acquisition

601 **Declaration of interest**

602 None to declare.

## 603 **Acknowledgements**

604 The authors would like to thank the experts who participated in the focus groups for  
605 the development of the assessments, interviewers and samplers who went to the  
606 farms to collect milk and information, as well as the French farmers who collaborated  
607 in the samplings and surveys. Authors would also like to acknowledge the help they  
608 received when drafting this article from the UMRH Scientific Writing Workshop  
609 chaired by Isabelle Veissier and Isabelle Ortigues-Marty. The authors also thank  
610 MetaForm Langues (Clermont-Ferrand, France) for checking and amending the  
611 English text.

## 612 **Financial support statement**

613 This research was supported by the Ministry of Agriculture in the framework of the  
614 CasDAR Qualenvic project (N°1251).

## 615 **References**

- 616 Agence française de sécurité sanitaire des aliments (AFSSA) 2009. Étude Individuelle  
617 Nationale des Consommations Alimentaires 2 (INCA 2) 2006-2007. AFSSA, Maisons-  
618 Alfort Cedex, France.
- 619 Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail  
620 (ANSES) 2011. Actualisation des apports nutritionnels conseillés pour les acides gras -  
621 Rapport d'expertise collective. ANSES, Maisons-Alfort Cedex, France.
- 622 Banks JM 2007. Cheese yield. In Cheese Problems Solved (ed. P.L.H. McSweeney), pp.  
623 100–114. Woodhead Publishing Limited, Cambridge, England.
- 624 Bockstaller C, Beauchet S, Manneville V, Amiaud B and Botreau R 2017. A tool to design  
625 fuzzy decision trees for sustainability assessment. Environmental Modelling & Software  
626 97, 130–144.

- 627 Botreau R, Beauchet S, Laurent C, Hulin S, Hérisset R, Thiollet-Scholtus M, Kanyarushoki C,  
628 Boucault P, Renaud-Gentié C and Jourjon F 2018. Une méthode pour évaluer  
629 conjointement performance environnementale et qualité globale des produits –  
630 Application aux filières vin et lait. *Innovations Agronomiques* 63, 23–42.
- 631 Botreau R, Capdeville J, Perny P and Veissier I 2008. Multicriteria evaluation of animal  
632 welfare at farm level: An application of MCDA methodologies. *Foundations of Computing  
633 and Decision Sciences* 33, 287–316.
- 634 Campolongo F, Cariboni J and Saltelli A 2007. An effective screening design for sensitivity  
635 analysis of large models. *Environmental Modelling & Software* 22, 1509–1518.
- 636 Cash D, Clark WC, Alcock F, Dickson NM, Eckley N and Jäger J 2002. *Saliency, Credibility,  
637 Legitimacy and Boundaries: Linking Research, Assessment and Decision Making.*  
638 Social Science Research Network, Rochester, NY, USA.
- 639 Chassaing C, Sibra C, Verbič J, Harstad OM, Golecký J, Martin B, Ferlay A, Constant I,  
640 Delavaud C, Hurtaud C, Pongrac VŽ and Agabriel C 2016. Mineral, vitamin A and fat  
641 composition of bulk milk related to European production conditions throughout the year.  
642 *Dairy Science & Technology* 96, 715–733.
- 643 Clark S and Bodyfelt FW (Eds) 2009. *The sensory evaluation of dairy products.* Springer,  
644 New York, NY, USA.
- 645 Coppa M, Ferlay A, Chassaing C, Agabriel C, Glasser F, Chilliard Y, Borreani G, Barcarolo  
646 R, Baars T, Kusche D, Harstad OM, Verbič J, Golecký J and Martin B 2013. Prediction  
647 of bulk milk fatty acid composition based on farming practices collected through on-  
648 farm surveys. *Journal of Dairy Science* 96, 4197–4211.
- 649 Coulon J-B 2008. Déterminants de la qualité des produits animaux. *Oléagineux, Corps gras,  
650 Lipides* 15, 126–133.
- 651 Coulon J-B, Delacroix-Buchet A, Martin B and Pirisi A 2004. Relationships between ruminant  
652 management and sensory characteristics of cheeses: a review. *Le Lait* 84, 221–241.

653 Ferlay A, Martin B, Pradel Ph, Coulon JB and Chilliard Y 2006. Influence of Grass-Based  
654 Diets on Milk Fatty Acid Composition and Milk Lipolytic System in Tarentaise and  
655 Montbéliarde Cow Breeds. *Journal of Dairy Science* 89, 4026–4041.

656 Frétin M, Martin B, Buchin S, Desserre B, Lavigne R, Tixier E, Cirié C, Bord C, Montel M-C,  
657 Delbès C and Ferlay A 2019. Milk fat composition modifies the texture and appearance  
658 of Cantal-type cheeses but not their flavor. *Journal of Dairy Science* 102, 1131–1143.

659 García V, Rovira S, Boutoial K and López MB 2014. Improvements in goat milk quality: A  
660 review. *Small Ruminant Research* 121, 51–57.

661 Gaucher I 2007. Caractéristiques de la micelle des caséines et stabilité des laits : de la  
662 collecte des laits crus au stockage des lait UHT. PhD Thesis, Agrocampus Ouest,  
663 Rennes, France.

664 Gaucher I, Tanguy G, Fauquant J, Jardin J, Rousseau F, Robert B, Madec M-N and  
665 Gaucheron F 2011. Proteolysis of casein micelles by *Pseudomonas fluorescens* CNRZ  
666 798 contributes to the destabilisation of UHT milk during its storage. *Dairy Science &*  
667 *Technology* 91, 413.

668 Guinee TP, Mulholland EO, Kelly J and Callaghan DJO 2007. Effect of Protein-to-Fat Ratio  
669 of Milk on the Composition, Manufacturing Efficiency, and Yield of Cheddar Cheese.  
670 *Journal of Dairy Science* 90, 110–123.

671 Halberg N, van der Werf HMG, Basset-Mens C, Dalgaard R and de Boer IJM 2005.  
672 Environmental assessment tools for the evaluation and improvement of European  
673 livestock production systems. *Livestock Production Science* 96, 33–50.

674 Harris J 1998. Raw milk grading using fuzzy logic. *International Journal of Dairy Technology*  
675 51, 52–56.

676 Haug A, Høstmark AT and Harstad OM 2007. Bovine milk in human nutrition – a review.  
677 *Lipids in Health and Disease* 6, 25.

678 Hurtaud C, Dutreuil M, Coppa M, Agabriel C and Martin B 2014. Characterization of milk  
679 from feeding systems based on herbage or corn silage with or without flaxseed and  
680 authentication through fatty acid profile. *Dairy Science & Technology* 94, 103–123.

681 Kindstedt PS 2014. The basics of cheesemaking. In Cheese and microbes (ed. Donnelly  
682 C.W.). ASM Press, Washington, DC, USA, pp. 17-38.

683 Lairez J, Feschet P, Aubin J, Bockstaller C and Bouvarel I 2015. Agriculture et  
684 développement durable. Guide pour l'évaluation multicritère. Éditions Quae, Paris,  
685 France.

686 Lebacqz T, Baret PV and Stilmant D 2013. Sustainability indicators for livestock farming. A  
687 review. *Agronomy for Sustainable Development* 33, 311–327.

688 Linstone H and Turoff M 1975. The Delphi Method: Techniques and Applications. Addison-  
689 Wesley Company, Reading, MA, USA.

690 Martin B, Graulet B, Uijtewaal A, Ferlay A, Coppa M and Rémond D 2019. Contribution des  
691 produits laitiers aux apports nutritionnels selon la nature des fourrages distribués aux  
692 vaches laitières. *Fourrages* 239, 193–202.

693 Morris MD 1991. Factorial Sampling Plans for Preliminary Computational Experiments.  
694 *Technometrics* 33, 161–174.

695 Müller-Lindenlauf M, Deittert C and Köpke U 2010. Assessment of environmental effects,  
696 animal welfare and milk quality among organic dairy farms. *Livestock Science* 128,  
697 140–148.

698 Prache S, Santé-Lhoutelier V, Adamiec C, Astruc T, Baeza-Campone E, Bouillot PE,  
699 Clinquart A, Feidt C, Fourat E, Gautron J, Guillier L, Kesse-Guyot E, Lebret B, Lefevre  
700 F, Martin B, Mirade PS, Pierre F, Remond D, Sans P, Souchon I, Girard A, Le Perchec  
701 S, Raulet M and Donnars C 2020. Qualité des aliments d'origine animale, Synthèse de  
702 l'expertise collective. INRAE, Paris, France.

703 Reif GD, Shahani KM, Vakil JR and Crowe LK 1976. Factors Affecting B-Complex Vitamin  
704 Content of Cottage Cheese. *Journal of Dairy Science* 59, 410–415.

705 Verdier-Metz I, Martin B, Pradel P, Albouy H, Hulin S, Montel M-C and Coulon J-B 2005.  
706 Effect of grass-silage vs. hay diet on the characteristics of cheese: interactions with the  
707 cheese model. *Le Lait* 85, 469–480.

708 Zahm F, Viaux P, Vilain L, Girardin P and Mouchet C 2008. Assessing farm sustainability  
709 with the IDEA method - From the concept of agriculture sustainability to case studies  
710 on farms. *Sustainable Development* 16, 271–281.

711 Zucali M, Battelli G, Battini M, Bava L, Decimo M, Mattiello S, Povolo M and Brasca M 2016.  
712 Multi-dimensional assessment and scoring system for dairy farms. *Italian Journal of*  
713 *Animal Science* 15, 492–503.

714

715 **Table 1**716 Sensitivity analysis of the UHT milk and cheese assessments of bovine bulk tank milk<sup>1</sup>

Indicators	Overall score		Nutritional dimension score		Health dimension score		Technological dimension score		Sensory dimension score	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
	UHT milk assessment									
Pseudomonas level	1.46	4.02	0.42	0.74	0.76	2.55	1.86	5.67	5.88	10.71
pH <sub>initial</sub>	1.41	2.19					6.90	10.79		
Calcium concentration	0.74	0.66	2.20	1.67						
Total $\beta$ -carotenes concentration	0.73	0.49			2.22	1.19				
Vitamin B2 concentration	0.67	0.32	1.88	0.47						
Lactose concentration	0.62	0.50			1.81	1.24				
Cheese assessment										
Pseudomonas level	1.55	2.82	1.08	2.15	0.52	1.00	0.88	1.79	2.54	5.22
Total caseins concentration	0.94	0.62	2.23	1.60	0.92	0.87	2.10	1.20		
Total bacteria count	0.82	1.48							1.89	3.17
ALA concentration	0.71	0.41	2.24	1.24	0.76	0.33				
Urea concentration	0.68	0.76					1.66	1.79	0.67	0.77
Calcium concentration	0.64	0.44	1.56	0.74			0.83	0.77		
Coliforms level	0.58	1.72							1.05	3.09

717 Abbreviations: UHT = Ultra-high temperature; ALA =  $\alpha$ -linolenic acid (C18:3 n-3)718 <sup>1</sup> Presentation of the seven most influential indicators of the overall score of each assessment. For the full sensitivity analysis, see Tables S3 and S4

719 **Table 2**

720 Mean year and season milk quality scores on 10 at global and dimensions level obtained on 30 dairy cow farms

Scores <sup>1</sup>	UHT milk assessment			Raw milk cheese assessment		
	Grazing	Indoor	Season effect <sup>2</sup>	Grazing	Indoor	Season effect <sup>2</sup>
Year overall quality	2.6 ± 0.5 (1.5-3.6)			3.5 ± 0.7 (2.1-4.8)		
Season overall quality	2.8 ± 0.5 (1.6-3.8)	2.1 ± 0.4 (1.3-2.9)	***	3.7 ± 1.0 (1.9-5.8)	3.1 ± 1.0 (1.1-5.1)	*
Dimensions of milk quality						
Sensory	7.2 ± 0.0	7.2 ± 0.0	-	4.2 ± 1.0 (2.2-5.8)	3.4 ± 0.8 (2.0-5.3)	**
Technological	4.3 ± 1.5 (1.2-6.5)	3.0 ± 1.5 (1.0-6.5)	***	4.4 ± 1.2 (2.5-7.4)	4.5 ± 1.5 (1.5-7.6)	ns
Health	2.8 ± 1.0 (1.2-5.6)	2.0 ± 0.9 (0.8-4.7)	**	6.7 ± 0.9 (4.6-8.1)	4.5 ± 1.3 (0.8-6.6)	***
Nutritional	3.0 ± 1.1 (1.3-5.2)	2.9 ± 1.3 (0.7-5.8)	ns	3.2 ± 1.3 (1.5-6.1)	3.4 ± 1.3 (1.1-6.1)	ns

721 Abbreviations: UHT = Ultra-high temperature; ns = not significant.

722 <sup>1</sup> Mean ± SD (Minimum-Maximum)723 <sup>2</sup> Symbols indicate a difference between grazing and indoor period's scores at P ≤ 0.05 (\*), P ≤ 0.01 (\*\*) and P ≤ 0.001 (\*\*\*) on the considered dimension  
724 and product [results from the Student's test or Wilcoxon test when data could not meet the normality assumption (sensory and health scores in cheese  
725 assessment and technological and health scores in milk assessment)].



726 **Figure captions**

727 **Fig. 1.** Construction process of the multicriteria assessment model from the  
728 participative approach

729 **Fig. 2.** Structure of the multicriteria assessment of the bovine intrinsic milk quality for  
730 ultra-high temperature (UHT) milk production per dimension: A) sensory,  
731 technological, health and B) nutritional quality. \* Ponderation of the dimension to the  
732 global quality score;  $\xrightarrow{100}$  Ponderation of the observation (indicators, subcriteria,  
733 criteria or principles) to the higher scale one (subcriteria, criteria, principles and  
734 dimensions , respectively); <sup>a</sup> CFU: Colony-forming unit; <sup>b</sup> FA: Fatty acids.

735 **Fig. 3.** Structure of the multicriteria assessment of the bovine intrinsic milk quality for  
736 cheese production per dimension: A) sensory, B) technological, C) health and  
737 nutritional quality. \* Ponderation of the dimension to the global quality score;  $\xrightarrow{100}$   
738 Ponderation of the observation (indicators, subcriteria, criteria or principles) to the  
739 higher scale one (subcriteria, criteria, principles and dimensions , respectively); <sup>a</sup>  
740 CFU: Colony-forming unit; <sup>b</sup> FA: Fatty acids; <sup>c</sup> ALA:  $\alpha$ -linolenic acid (C18:3 n-3); <sup>d</sup> LA:  
741 linoleic acid (C18:2 n-6).

742

## ① Definition of the scope **L**

- A. Aim of the assessment and future users
- B. System to evaluate  
description of the object, spatial and temporal scales
- C. Prerequisites  
financial and time costs, ability to run it



## ② Definition of the conceptual and methodological frameworks

### A. Conceptual framework: Hierarchical structure

Construction of the structure from overall quality to indicators:

Definition of dimensions **L**, principles, criteria and indicators **S**

Simplification of the structure **L S** (choice of indicators)

### B. Methodological framework: Evaluation

Interpretation of indicators **S** (threshold values)

Agregation of scores (weighting of dimensions **L**, of other levels **S**, compensation **L**)



## ③ Validation of the assessment **L**

### A. Sensitive analysis

Overview of the assessment's behaviour

Identification of influential indicators

### B. Test on 30 dairy farms

Overview of the assessment's discrimination rate

Range and dispersion of the farms' scores

**L** Leader experts group

**S** Specific experts group









