

1 **Individual heterogeneity as a pitfall in population estimates based on non-invasive**
2 **genetic sampling – review and recommendations**

3 Cornelia Ebert, Felix Knauer, Ilse Storch, Ulf Hohmann

4

5 Cornelia Ebert & Ulf Hohmann: Department of Forest Ecology and Wildlife, Research
6 Institute for Forestry and Forest Ecology of Rhineland-Palatinate, Hauptstrasse 16, D-67705
7 Trippstadt, Germany - e-mail addresses: cebert@gmx.de (Cornelia Ebert),
8 ulf.hohmann@wald.rlp.de (Ulf Hohmann) - telephone numbers: 0049-631-3117110 (Cornelia
9 Ebert), 0049-6306-911148 (Ulf Hohmann) - fax number: 0049-6306-911200

10 Felix Knauer: Department Wildlife Ecology and Management, University of Freiburg and
11 Reseach Institute for Wildlife Ecology, present address: Savoyenstr. 1, A-1160 Vienna,
12 felix.knauer@fiwi.at, telephone number: 0043-1-4890915-184, fax number: 0043-1-4890915-
13 733

14 Ilse Storch: Department Wildlife Ecology and Management, University of Freiburg,
15 Tennenbacher Strasse 4, D-79106 Freiburg, Germany - e-mail address:
16 ilse.storch@wildlife.uni-freiburg.de telephone number: 0049-761-2033797 - fax number:
17 0049-761-2033667

18

19 Corresponding author: Cornelia Ebert

20

21 **Keywords: capture-recapture; genotyping; hair sampling; individual heterogeneity;**
22 **population estimate; faeces; wildlife management**

23

24 **Abstract:**

25 In recent years much progress has been made in non-invasive genetic methods for various
26 purposes including population estimation. Previous research focused on optimizing laboratory

27 protocols and assessing genotyping errors. However, an important source of bias in
28 population estimates still remains in the field sampling methods. The probability of animals
29 being sampled can vary due to sex, age, social status or home range location. In this paper,
30 relevant literature is reviewed to provide an overview of the occurrence of individual
31 heterogeneity (IH) in the field and how it can be minimised, e.g. by adaptation of sampling
32 design. Thirty-eight articles describing non-invasive population estimation for 12 mammal
33 and two bird species were surveyed. The majority of studies discussed IH as a potential
34 problem. The detectability of IH via goodness-of-fit testing depended on the average capture
35 probability reported in the studies. Field tests for assessing variation in sampling probabilities
36 or validating estimations were carried out in only 11 out of the 38 studies. Results of these
37 tests point out that IH is a widespread problem in non-invasive population estimation which
38 deserves closer attention not only in the development of laboratory protocols, but also
39 concerning the sampled species' characteristics and the field methods. IH can be reduced in
40 the field by carefully adapting the sampling design to the characteristics of the studied
41 population. If this is not reasonable, it may be better to switch to a different sampling strategy.

42

43 **Introduction:**

44 Reliable estimation of population size remains a major challenge in wildlife research and
45 management. In recent years non-invasive DNA-based population estimation methods have
46 been widely applied in a variety of species. Several standard approaches have been modified
47 to fit genetical implementation, among these are rarefaction (e.g. Frantz et al. 2004) and
48 capture-recapture (e.g. Boulanger & McLellan 2001). In their conventional form, both
49 methods presuppose capture or killing of animals or rely on direct sightings, and are
50 challenged by the possibility of heterogeneous detection probabilities amongst the studied
51 population (Borchers et al. 2002, Petit & Valière 2006). Being most frequently used, capture-
52 recapture (CR) methods are especially vulnerable with respect to individual heterogeneity

53 (Pledger & Efford 1998, Link 2004, Lukacs & Burnham 2005), i.e. differences between
54 individuals of a population in the probability of being captured (Borchers et al. 2002).
55 Capture- and recapture probabilities may be influenced by age, sex, social status, and
56 individual experience (Baber & Coblenz 1986, Piggott & Taylor 2003). This can generate
57 severe bias in population estimates (White et al. 1982, Minta & Mangel 1989, Sweitzer et al.
58 2000). Individual heterogeneity (IH) can be accounted for with different modelling
59 approaches (see e.g. Otis et al. 1978, Chao 1987, Chao & Jeng 1992, Pledger & Efford 1998).
60 But the power of goodness-of-fit (GOF) tests and model selection procedures to detect IH in a
61 given data set often is low (Menkens & Anderson 1988, McKelvey & Pearson 2001).
62 Furthermore, as Link (2003 and 2004) has stated recently, IH is far more difficult to model
63 than has previously been recognized, modelling being especially problematic if the causes and
64 extent of IH are unknown. Thus, in order to allow accurate population estimates, IH should
65 either be minimised or quantified as far as possible (Petit & Valière 2006).

66 Methods based on non-invasive genetic sampling offer solutions for estimation of population
67 size without capturing or killing animals, making them advantageous for rare or endangered
68 species (Kohn et al. 1999, Taberlet et al. 1999, Mills et al. 2000, Piggott & Taylor 2003). It
69 was suggested that the absence of handling can overcome the effects of previous capture
70 history on subsequent catchability, thus certain sources of IH could be reduced (McKelvey &
71 Schwartz 2004, Petit & Valière 2006). The most commonly used non-invasive DNA sources
72 are hairs and faeces for mammals, as well as feathers and faeces for birds (Lukacs & Burnham
73 2005). Non-invasive methods have made CR approaches – which in their conventional form
74 are more suitable for small and abundant mammals – applicable for large, elusive and/ or
75 endangered mammal and bird species (Obbard et al. 2010).

76 However, despite their advantages, non-invasive genetic methods are also prone to
77 heterogeneity related to biological variability among individuals (Kohn et al. 1999, Wilson et
78 al. 2003, Boulanger et al. 2004b). Moreover, in non-invasive methods IH can interact with

79 bias caused by genotyping errors. Allelic dropout and false alleles can create ‘new’ false
80 individuals, leading to overestimation in population estimates because recaptures may be
81 concealed, resulting in a decreased recapture rate (Creel et al. 2003, McKelvey & Schwartz
82 2004). Furthermore, there are some issues in non-invasive genetic CR which are not
83 problematic in conventional CR. In genetic CR the total number of marks in the population is
84 not known and marks may not be unique, because only a subset of each animal’s genome is
85 used for identification (Lukacs and Burnham 2005). Therefore, the danger of
86 misidentification is increased compared to conventional CR. Also, a ‘sampling occasion’ can
87 be more difficult to define compared to a ‘capture occasion’, because the moment of the
88 deposition of a sample – e.g. hair or faeces – can not be assessed precisely. This can
89 compromise the concept of population closure (Lukacs & Burnham 2005). Thus, despite the
90 high potential of non-invasive genetic techniques, there are several issues which can
91 complicate the application of a CR framework for population estimation - in addition to the
92 difficulties already present in the conventional approach.

93 Until now, great progress has been made in genetic techniques. In particular, much effort has
94 been devoted to quantifying and reducing genotyping errors (Taberlet et al. 1999, Paetkau
95 2003, Broquet & and Petit 2004, Roon et al. 2005, Miquel 2006). In contrast, fewer attempts
96 have been made for assessing the extent and causes of IH in the field – i.e. due to biological
97 characteristics of the sampled species, to individual attributes or due to sampling procedures
98 (Boulanger et al. 2006). However, information about the causes and the extent of IH is
99 essential to improve sampling designs (Boulanger et al. 2004a). Furthermore, IH in
100 combination with uncertainties caused by genotyping errors can cause multiplicative effects
101 and thus lead to an increase in overall bias. Therefore, it is crucial to address both – IH and
102 genotyping problems – very carefully in order to minimise bias in population estimates.

103 Based on the recent peer-reviewed literature, this paper aims at

- 104 1) providing a survey of the occurrence and treatment of IH in non-invasive population
105 estimation studies, especially with respect to different sampling strategies
- 106 2) assessing the impact of sample size and capture probability (p) on the detectability of
107 IH via GOF tests or model selection procedures in CR studies
- 108 3) comparing different methods which seem suitable to assess IH in the field – also with
109 respect to the study species and its characteristics

110

111 **Material and Methods:**

112 This review is based on population genetic studies that involve non-invasive sampling for the
113 purpose of population estimation in mammals and birds. We performed a search in Swiss
114 Wildlife Information Service (SWISS) database for peer-reviewed publications using the
115 following search terms: hair trap, non-invasive sampling, genotyping, population estimates,
116 faeces sampling, hair sampling, genetic monitoring. The search yielded 104 titles and was
117 supplemented with published lists of references. In total, we detected 142 articles of which we
118 focused on 38 studies (complete list of references available on request from the corresponding
119 author). We only included papers in which the non-invasive sampling was *de facto* conducted
120 in the field and applied for population estimation; literature reviews and articles dealing with
121 single aspects in the development of sampling methods were excluded. We focused on studies
122 using hair, faeces and feathers as those are the main sources of non-invasive tissue samples.
123 Other sources (e.g. urine, shed skin or buccal swabs) have been much less employed for
124 population estimation until now (Broquet et al. 2007). We also included cases in which a
125 combination of different sampling strategies was applied. We restricted our review to studies
126 using CR or rarefaction (also termed ‘accumulation curve’ methods; see e.g. Kohn et al. 1999,
127 Eggert et al. 2003) approaches for population estimation, as those are the most commonly
128 used and more prone to be biased by IH compared to e.g. estimation of minimum densities or
129 minimum number alive.

130 For each study we assessed if IH had been mentioned, i.e. considered as a factor potentially
131 influencing the population estimate. Additionally, we recorded if IH was detected – e.g. via
132 likelihood ratio tests (program CAPWIRE, Miller et al. 2005), Chi²- tests (program
133 CAPTURE, Otis et al. 1978), via Akaike’s Information Criterion (AIC, e.g. in program
134 MARK, White et al. 1999) or GOF testing in program U-CARE (Choquet et al. 2005).
135 Furthermore, IH can be discerned in uneven ‘capture frequencies’ of sampled individuals
136 (Kohn 1999, Scheppers et al. 2007). The power of tests to detect IH can depend on capture
137 probability (p) (Pollock et al. 1990, Boulanger et al. 2002). Additionally, we suspected the
138 number of sampling occasions and coverage, i.e. proportion of the population sampled, to
139 have an effect on IH detectability. The estimated coverage is significantly correlated with p
140 and was included because not all reviewed studies provided estimates of p. We used logistic
141 regression to evaluate the impact of p, coverage, and sampling occasions on the probability of
142 IH being detected. In this context, we evaluated studies in which IH was detected in the
143 capture frequency or via field test, but not in the GOF or model selection tests as ‘not
144 detected’. We included also squared terms of p and coverage since data suggested an
145 optimum somewhere in between of the extreme values. Models were selected based on AIC.
146 For the logistic regression, we used every single population estimate reported in the 38
147 reviewed articles, since results of several study years or different study areas were often
148 included in one article, resulting in different p and population estimates. Since p and coverage
149 of different analyses reported within the same study could be correlated, we included the
150 studies as random factor. This worked well with coverage but not with p, because for p
151 sample size was low (only 39 of 76 analyses included estimates of p average) and the number
152 of studies reporting only one analysis was high. Therefore, in the case of p we averaged the
153 values for each study and conducted both, a weighted and an unweighted logistic regression
154 without random effects. All analyses were performed using program R (Ihaka and Gentleman

155 1996). For the mixed effect logistic regression model we used the function lmer of the
156 package lme4 (Bates and Maechler 2010).

157 For studies using CR methodology, we additionally recorded if IH was included in the
158 estimation model (which is not possible for rarefaction methods). We also searched for
159 studies in which field tests had been carried out parallel to the non-invasive sampling for
160 validation purposes. We put our special attention to methods and results of these studies and
161 aimed to assess if the applied field tests hold the potential to reveal IH.

162

163 **Results:**

164 The papers we reviewed dealt with 14 different study species, 12 mammals and 2 birds. In 30
165 of the 38 studies we included in our review, CR was the sole method applied to estimate
166 population size. Four studies used rarefaction analysis only and four used both methods
167 (Table 1). Hair was the DNA source in 22 of 38 studies altogether. In one of these cases, the
168 hair sampling was combined with harvest data and in another, faecal sampling was carried out
169 simultaneously. The remaining 16 studies relied on faeces as DNA source, one of them in
170 combination with feathers. IH bias was mentioned as a potential problem in 34 (89 %) of all
171 contributions, whereas it was modelled in 26 of the 34 capture-recapture studies (i.e. 76 %).
172 In 18 of the 34 capture-recapture articles, IH was detected via Chi²- tests, likelihood ratio tests
173 or with the help of AIC. A further 5 studies performed GOF tests but did not detect IH in their
174 data sets. In 6 of all 38 studies, tests either failed to detect IH or were not performed, but it
175 was nevertheless visible in the 'capture frequencies' (Table 1). Altogether, in 24 of the 38
176 studies (63 %) the data revealed the occurrence of IH amongst the studied population
177 independent of further field tests.

178

179 Assessment of IH via GOF tests and model selection procedures

180 In this section we are taking into account every single population estimate ($N = 76$) reported
181 in the 38 reviewed articles. More than half (53.8 %) of the 39 reported estimates of p average
182 lie below 0.2 (Fig. 1) and thus around or below the minimum recommended by Otis et al.
183 (1978) for reliable model selection and population estimates (Otis et al. recommend $p \geq 0.2$
184 for a population of 200 animals and state that p should never lie below 0.1. Recommended
185 minimum number of capture occasions is 5, better 7 to 10). Furthermore, IH was only
186 detected for p between 0.16 and 0.4. The proportion of population estimates in which IH was
187 detected increased with increasing p until $p = 0.4$ (Fig. 1). In none of 3 studies with $p \geq 0.4$ IH
188 was detected. However, in one of those studies sample size was too small to carry out tests in
189 program CAPTURE (see table 1, Belant et al. 2005, sampling Sand Island). The logistic
190 regression showed no impact of coverage on the detectability of IH (Table 2). In the case of p
191 and p^2 , the results suggest that there is an effect on the detectability of IH (Table 3 and 4). IH
192 detectability is highest at p values around 0.3 (Fig 2). The most supported model does not
193 include the number of sampling occasions, but a model including sampling occasions is
194 ranked marginally below ($\Delta AIC < 2$), indicating a potential influence (Burnham and
195 Anderson 1998). It seems possible that with increasing number of sampling occasions, the
196 detectability of IH increases. Models including an interaction between p and the number of
197 sampling occasions were not supported (table 3).

198

199 Assessment of IH via field tests

200 Field tests suitable for assessing the occurrence of IH bias were performed in 11 of the 38
201 studies. In 7 of these 11 cases, IH was actually detected (Table 1). In one of the 4 other cases,
202 IH was found to be present in the hair sampling part of the study, but was strongly reduced by
203 sampling harvested animals as an additional strategy (Dreher et al. 2007). Furthermore, in 8
204 of the 11 studies, IH was detected via GOF testing or in the 'capture frequencies'. Thus, in
205 two cases where the field tests did not reveal IH, it seemed nevertheless to be present and

206 detectable in the data set. Furthermore, in two cases IH was detected through the field test but
207 not in the data.

208

209 Field tests in detail

210 Using radiotelemetry Kohn et al. (1999) found IH to be present in their population under
211 study: 12 radio-collared coyotes (*Canis latrans*) made use of the area to different degrees, and
212 the number of faeces deposited correlated with their relative use of the study area. This IH
213 was also reflected in the 'capture frequencies' of the sampled individuals. Faeces sampling of
214 a coyote population in Central Alaska exhibited IH with respect to age and home range as
215 well as resident status. This was revealed through radiotelemetry of 15 collared adult resident
216 individuals which showed higher survival and recapture rates compared to juveniles and
217 transient or edge individuals (Prugh et al. 2005). Furthermore, the model selection process in
218 program MARK detected IH in the data. This was also the case in a hair sampling study on
219 grizzly bears (*Ursus arctos*): Boulanger et al. (2004a) used location data of 12 GPS-collared
220 bears to evaluate potential bias. They found p to be greater for males than for females and also
221 to be influenced by capture history (i.e. differences between collared and non-collared
222 individuals). The latter was also detected in another study (Boulanger et al. 2004b), in which
223 radiotelemetry was conducted over three years on a total of 35 bears and compared to hair
224 sampling data collected in the same area. Additionally the p of females with cubs differed
225 from those of the rest of the population. In a study carried out by Wasser et al. (2004), grizzly
226 bear faeces sampling data was compared to hair sampling and radiotelemetry data collected
227 simultaneously in the same area. Faeces collection was conducted with the help of trained
228 dogs. This seemed to be an effective and relatively less biased method compared to hair
229 sampling at baited stations. In the latter, close kin (i.e. females with their offspring) were
230 considerably less represented. However, the sample size of matched faeces and telemetry data
231 was too small to allow more fine-grained comparisons, and IH was neither detected in the

232 field test nor in the data set (Wasser et al. 2004). In a black bear (*Ursus americanus*) hair
233 sampling study accompanied by radiotelemetry, Dreher et al. (2007) used harvested bears as
234 an additional sample. Due to this combination, the IH, which would have been present if hair
235 sampling or harvest data were used alone, was strongly reduced (B. Dreher, pers. comm.) and
236 thus not detected in the field test and in the data set.

237 Wilson et al. (2003) carried out a faeces sampling study on badgers (*Meles meles*) in which
238 they used video control of a largely marked population to validate their rarefaction estimate.
239 They succeeded in sampling almost the entire population by collecting faeces at latrines near
240 badger setts and did not detect sex- or age class-bias. However, considerable variation existed
241 in the numbers of samples obtained from the different individuals. Moreover some known
242 individuals never used the sampled latrines and thus were not identified via faeces sampling.
243 These results suggest the incidence of IH - e.g. due to variation in individual behaviour -
244 which in that case might not have compromised the estimation, because such a high
245 proportion of the population was actually sampled. A hair sampling study conducted by
246 Scheppers et al. (2007) at badger setts - simultaneously surveyed via direct observations -
247 yielded similar results. The ease with which badgers were sampled varied considerably
248 between setts; hair traps were not visited equally often by all members of the groups. Using
249 baited hair traps and applying direct observation as a validation method, the results of Frantz
250 et al. (2004) show a comparable pattern. Even though obvious variation in the individual
251 sampling frequency existed, almost all badgers present in the area were sampled. Thus, the
252 rarefaction analysis yielded quite reliable results. However, the IH observed in the three
253 badger studies might have been crucial in other populations or situations, e.g. when a lower
254 proportion of the population is represented in the samples, especially when CR methods are
255 applied (Pollock 1982). Furthermore, video control as well as direct observation in all three
256 studies focused on obtaining an independent census of the sampled badger groups, not on
257 observing the sampling behaviour itself. As a consequence potential sources of IH like e.g.

258 dependence of latrine use or access to bait on social status may remain undetected. The same
259 seems to hold true for a study on the lesser horseshoe bat (*Rhinolophus hipposideros*): the
260 non-invasive population estimate was validated via direct counts of bats in their day roosts
261 (Puechmaille & Petit 2007). The direct counts did not reveal any IH in the faeces sampling.
262 However, IH was detected in the sampling data of several of the sampled bat colonies via
263 likelihood ratio- or simulation tests.

264 In a faeces sampling study on wolves (*Canis lupus*) in the Italian Alps, Marucco et al. (2009)
265 used an evaluation system for age-dependent marking behaviour related to defecation. Due to
266 the fact that part of the population was radiocollared or otherwise known, it was possible to
267 discriminate between faeces deposited by adult wolves for marking purposes and ‘non-
268 marking’ faeces. The authors detected age- and status- dependent IH in defecation behaviour
269 and concluded that they would have missed a considerable part of the juvenile population if
270 they had not adapted their search pattern. However, by means of the field tests, Marucco et al.
271 (2009) were able to apply and confirm a representative faeces sampling strategy.

272

273 **Discussion:**

274 Most researchers seem to be aware of IH being a major problem present in population
275 estimation based on non-invasive genetic methods. The vast majority of articles dealing with
276 non-invasive methods applied for such purpose mention or discuss this problem. In most
277 studies based on CR approaches, the authors attempted to account for potential bias by
278 employing models which incorporate IH (Chao 1987). However, as long as the different
279 sources and the extent of IH are unknown, the results of population estimations are strongly
280 model dependent and might not reflect reality (Link 2003, 2004). Furthermore, the different
281 methods to test for IH in the data set may have limited power and thus often fail to detect IH
282 (Boulanger & McLellan 2001, Miller et al. 2005). In the literature, it has been mentioned that
283 the power of such test procedures is especially low for low p (Menkens & Anderson 1988,

284 Boulanger & McLellan 2001). The results of our analysis support this finding: they indicate
285 an impact of p on the detectability of IH via GOF tests and model selection procedures. This
286 effect does not seem to depend on the type of test and the software used (logistic regression
287 with test type as additional covariate showed no significant effect; results not shown). In our
288 analysis, we used a very conservative approach by averaging over the studies and using both,
289 weighted and unweighted values. Both approaches show very similar results, indicating the
290 results being robust to details in the analysis. In studies with low p , IH was detected
291 considerably less often compared to studies with higher p . The highest proportion of detected
292 IH was attributed to studies with p between 0.2 and below 0.4. Interestingly, in none of the
293 three studies reporting p higher than 0.4, tests suggested the incidence of IH. This might be
294 due to the fact that IH bias becomes much less problematic – perhaps even negligible – when
295 p is high, which has been shown in simulation studies (J. Boulanger, pers. comm.). When
296 most animals in a population are actually captured or sampled, the differences in p between
297 individuals have much less impact on the population estimate (Pollock 1990, Lukacs &
298 Burnham 2005). Thus, IH might not be reflected in GOF testing or model selection when p is
299 high. Even though the number of sampling occasions for each study was not included in the
300 most supported model, there seems to be an indication for a certain influence on the
301 detectability of IH, because the model including sampling occasions was ranked only
302 marginally inferior to the best model. The more sampling occasions are carried out, the better
303 might be the ability to detect IH via GOF testing. However, this point needs further
304 investigation before a clear conclusion on sampling occasions can be drawn. In contrast to p ,
305 an impact of coverage on the detectability of IH was not supported in our analysis, despite the
306 correlation of coverage and p .

307 It should be mentioned that with our analysis we are not able to distinguish if a negative result
308 of testing for IH is due to lack of power and test failure or because there simply is no IH
309 present in a given data set. However, regarding the existing literature including simulation

310 studies and studies on populations of known size, IH seems to be almost ubiquitous in non-
311 invasive sampling data sets like in conventional CR (Pollock 1990, Borchers et al. 2002,
312 Knapp et al. 2007, Lukacs & Burnham 2005). Therefore, it seems much more likely that a
313 negative test result is caused by low test power than that a data set is really homogeneous,
314 particularly when p is low. In recent years, new modelling approaches, e.g. multistate and
315 multievent models (Pradel 2005), have been developed which might allow a more flexible
316 handling of CR data in presence of IH (Crespin et al. 2008). In this context GOF testing using
317 nonparametric methods, like in program U-CARE, seems to be quite promising compared to
318 conventional methods (Choquet et al. 2009, Cubaynes et al. 2010).

319 In 5 of the articles reviewed here, IH was not detected by the data tests which were
320 performed. However, in 2 of the 5 cases, additional field tests were carried out, and both of
321 them revealed IH. In general, data tests and/or pronounced differences in the ‘capture
322 frequencies’ indicate presence of IH bias without carrying out extra field tests, but often
323 further investigations would be required to uncover the causes of IH. Since many different IH
324 sources exist, they can influence estimations in different ways and this effect may also depend
325 on the sampling design (Crespin et al. 2008). Models that are relatively robust to IH generally
326 show reduced precision of estimate (Boulangier 2004b). This may not be tolerable in cases
327 where an accurate estimate is particularly important, e.g. when the spread of diseases is
328 concerned (Artois et al. 2002) or when management plans for rare or endangered species are
329 considered (Guschanski et al. 2009). However, for endangered species, overestimating a
330 population is much more critical than underestimating it (Meijer et al. 2008), so some
331 underestimation bias may be tolerable in certain cases.

332

333 Assessment of IH via field tests

334 The choice of methods to test for IH in the field strongly depends on the observed species and
335 its behavioural patterns as well as its space and habitat use. Thus, e.g. for badgers, which live

336 in social groups, share setts and make rather small-scale movements, video control or direct
337 observations at setts seem to be an adequate method to validate non-invasively obtained
338 estimates (see e.g. Frantz et al. 2004; Scheppers et al. 2007). Contrastingly, for highly mobile
339 species like bears and also coyotes, radiotelemetry may be more promising. The suitability of
340 a field method to test for IH furthermore depends on the applied sampling strategy. For
341 example, video control or direct observations can be appropriate for surveillance of discrete
342 sampling stations like hair traps or badger setts, but will not be suitable for large-scale
343 sampling designs like e.g. line transects. Radiotelemetry may be more effective to observe
344 movements and transect- or trap-encounter rates of animals on a large scale. Furthermore,
345 radiotelemetry is useful for obtaining information on spatial distribution and home range sizes
346 of animals in order to fit sampling designs and to account for closure violations and edge
347 effects (Boulanger et al. 2004b, Dreher et al. 2007). The feasibility of a sampling method for a
348 given species or population can depend on spatial characteristics like home range sizes and
349 distribution of animals in the sampled area. Settlage et al. (2008) found hair sampling of black
350 bears via baited sampling stations impractical for the Southern Appalachian region: due to
351 small home range sizes of the resident bears, sampling probabilities were low and biased. In
352 order to yield a reliable estimate, a much higher sampling intensity would have been
353 necessary (Settlage et al. 2008). Grizzly bears showed considerably higher p with comparable
354 sampling intensities because of their larger average home ranges (Boulanger et al. 2004b,
355 McLoughlin et al. 2003).

356

357 Interactions between sampling strategies and study species' characteristics

358 The occurrence and/ or extent of IH may differ dependent on the applied sampling method.
359 'Active' sampling methods, like hair sampling via baited hair traps, presuppose that animals
360 actively approach the sampling station. In many species, it has been shown that individuals
361 show consistent or context-specific personality traits, e.g. they differ in their exploration

362 behaviour and their reactions towards newly introduced factors, which may affect their
363 sampling probability (Coleman and Wilson 1996, Ruis et al. 2000, Dingemanse et al. 2003,
364 Mettke-Hofmann et al. 2005). Furthermore individual experience and life history may
365 influence behaviour with respect to sampling stations. This could cause IH which is not
366 necessarily related to sex, age or social status, and which might be hard to quantify and very
367 difficult to account for in a model. Thus, in some cases, it may be reasonable to apply a
368 different sampling method. In this context, ‘passive’ sampling strategies like e.g. faeces
369 sampling along transects represent an alternative which may be less affected by individual
370 behaviour or status differences. This may hold true particularly for group living species:
371 interactions between animals can increase IH, especially when sampling concentrates on
372 defined stations like e.g. hair traps which require active approach. As an example, we
373 conducted a hair sampling pilot study on wild boar (*Sus scrofa*). Video observation at baited
374 hair traps revealed significant behavioural differences depending on age of the animals and on
375 their group status (Ebert et al. 2009). However, also for bears which can be considered as
376 living mainly solitary, it has been shown that via faeces sampling – a ‘passive’ sampling
377 strategy – a larger part of a population can be observed compared to hair sampling as an
378 ‘active’ approach (Wasser et al. 2004). Wasser et al. applied both methods in the same study
379 area and time period, and via hair sampling, only 46% of the individuals that were identified
380 via faeces sampling were detected. ‘Passive’ sampling methods in most cases will not yield
381 completely unbiased results (in fact, most of the faeces sampling studies reviewed here
382 reported IH in their data sets). Nevertheless, ‘passive’ sampling might rule out certain sources
383 of IH which are not avoidable in ‘active’ approaches and thus holds the potential to yield
384 results with smaller overall bias. However, in some (especially social and/ or territorial)
385 species, status or age differences between individuals may cause differences in faeces
386 deposition patterns, leading to IH in detection probabilities. This has been shown e.g. for
387 wolves (Marucco et al. 2009, Cubaynes et al. 2010). Thus, ‘passive’ sampling will not be

388 suitable in all cases, and at any rate the appropriate sampling strategy and design have to be
389 carefully tested for each particular species and population. Furthermore, the DNA quality of
390 faeces in some cases has been shown to be inferior to that of hair, thus population estimates
391 derived from faeces sampling data may be more in danger of bias due to genotyping errors
392 (Piggott and Taylor 2003).

393

394 **Recommendations:**

395 *Perform a pilot study - not only in the lab, but also in the field:* In any case, it is most
396 advisable that researchers who plan to establish population estimation based on non-invasive
397 genetic sampling perform pilot studies not only to assess genotyping error rates, but also to
398 detect sources of IH bias in the field. The fact that the majority of reviewed studies in which
399 such field tests were performed actually detected IH highly supports this recommendation.
400 The appropriate methods to assess IH in the field depend on the species or population under
401 study as well as on the applied sampling method.

402 *Do not rely solely on GOF testing and model selection procedures:* This holds especially true
403 when p and coverage are low! It can be reasonable to incorporate heterogeneity in an
404 estimation model even if tests suggest that there is no IH present in a data set, because their
405 power is often low. It is always recommended to include biological knowledge and
406 information about study species, habitat etc. to validate model choice.

407 *Try to reduce IH by adapting sampling design:* Knowledge about the sources and extent of IH
408 can enable researchers to adapt the sampling design to account for the bias. Among the
409 methods to reduce IH bias in the field, the application of two or more different sampling
410 strategies in combination seems especially promising (Dreher et al. 2007, Boulanger et al.
411 2008, Settlage et al. 2008). If multiple methods are used simultaneously to sample a
412 population, the impact of IH caused by any single method can be minimised (Pollock 1982,
413 Williams et al. 2002). The improvement of estimations based on multiple approaches

414 increases with decreasing correlation between the applied sampling methods (Boulanger et al.
415 2008). In the case of hair sampling, the use of unbaited sampling stations (e.g. installed at
416 trails or rubbing trees) and changing of sampling locations between sessions may be applied
417 to reduce IH due to competition between individuals for resources and due to “trap happy”
418 individuals (Scheppers et al. 2007, Boulanger et al. 2008). Collection of faeces samples with
419 the help of trained dogs seemed to increase detection rate and efficiency of the method
420 considerably, allowing a relatively representative and unbiased population survey compared
421 e.g. to hair sampling (Wasser et al. 2004, Long et al. 2007). Furthermore, it is advisable to
422 perform a sufficiently high number of sampling occasions in order to increase the overall
423 sampling probability and thus to facilitate accurate estimates.

424 *Try to sample a large part of the population:* As shown e.g. in the three badger studies
425 reviewed here, one effective way to reduce bias caused by IH is to sample a large proportion
426 of the population (Lukacs & Burnham 2005). This is generally desirable and has been
427 recommended in relevant literature many times before (see e.g. Otis et al. 1978, Pollock et al.
428 1982), but is certainly not always feasible. Furthermore, an increase in sample size can have
429 an unfavourable impact on non-invasive genetic population estimates: the more samples are
430 analysed, the higher the misidentification rate due to genotyping errors (McKelvey &
431 Schwartz). Thus, careful error-checking protocols for genotyping are crucial and genotyping
432 error rates should be determined in order to avoid an increase in bias through
433 misidentification (Maudet et al. 2004, Roon et al. 2005).

434 *Consider switching to other sampling strategies:* Adapting the sampling design may not
435 always be possible or may yield no success. Furthermore, an unsolved problem still remains:
436 even though detection of IH and its sources may be possible with methods like e.g.
437 radiotelemetry or video observation, the exact quantification of such variation and thus its
438 incorporation in estimation models still seems to be very difficult. Consequently, when
439 reduction and/ or modelling of IH is not possible, it can be recommendable to apply a

440 different sampling method in some cases. The suitability of a method can depend e.g. on
441 characteristics of the studied species, population, or study area. In some cases, ‘passive’
442 sampling approaches may yield more representative results compared to ‘active’ methods. In
443 case IH can not be reduced or avoided, a study should be designed in a way that it results in
444 capture probabilities between 0.2 and 0.4 to have an ample chance to detect existing IH.

445

446 In conjunction with problems caused by genotyping errors, IH is a highly challenging issue in
447 non-invasive population estimation. It is a well-known and explicitly discussed problem at
448 least with regard to its theoretical and model-based aspects. IH can be identified and strongly
449 reduced, when field sampling design and analytical approach are carefully prepared.
450 However, more attention should be given to the evaluation of field methods to bring forward
451 more effective and sustainable population estimates, which is especially important for
452 conservation of endangered species and even more in fragmented habitats.

453

454 **Acknowledgements** - we wish to thank J. Boulanger for giving valuable supplemental ideas
455 and J. Arnold, J. Hofmann, W. Maurer, K. Jochum as well as two anonymous reviewers, who
456 provided further helpful comments on earlier drafts of this manuscript. This study was granted
457 by the foundation “Rheinland-Pfalz fuer Innovation” and the Ministry for Environment,
458 Forestry and Consumer Protection in Rhineland-Palatinate, Germany. C. Ebert gratefully
459 acknowledges financial support from the FAZIT foundation.

460

461 **References:**

462 Banks, S.C., Piggott, M.P., Hansen, B.D., Robinson, N.A., Taylor, A.C. 2002: Wombat
463 coprogenetics: enumerating a common wombat population by microsatellite analysis of faecal
464 DNA.- Australian Journal of Zoology 59: 193-214.

465 Banks, S.C., Hoyle, S.D., Horsup, A., Sunnucks, P. & Taylor, A.C. 2003: Demographic
466 monitoring of an entire species (the northern hairy-nosed wombat, *Lasiorhinus kreftii*) by
467 genetic analysis of non-invasively collected material.- *Animal Conservation* 6: 101-107.

468 Bates, D. & Maechler, M. 2010: lme4: Linear mixed-effects models using S4 classes. R
469 package version 0.999375-33. <http://CRAN.R-project.org/package=lme4>.

470 Belant, J.L., Van Stappen, J.F. & Paetkau, D. 2005: American black bear population size and
471 genetic diversity at Apostle Islands National Lakeshore.- *Ursus* 16: 85-92.

472 Bellemain, E., Swenson, J.E., Tallmon, D., Brunberg, S. & Taberlet, P. 2005: Estimating
473 population size of elusive animals with DNA from hunter-collected feces: four methods for
474 brown bears.- *Conservation Biology* 19: 150-161.

475 Bellemain, E., Nawaz, M.A., Valentini, A., Swenson, J.E. & Taberlet, P. 2007: Genetic
476 tracking of brown bear in northern Pakistan and implications for conservation.- *Biological*
477 *Conservation* 134: 537-547.

478 Boersen, M.R., Clark, J.D. & King, T.L. 2003: Estimating black bear population density and
479 genetic diversity at Tensas River, Louisiana using microsatellite DNA markers.- *Wildlife*
480 *Society Bulletin* 31: 197-207.

481 Borchers, D.L., Buckland, S.T. & Zucchini, W. 2002: Estimating Animal Abundance –
482 Closed Populations.- *Statistics for Biology and Health*, Springer, London, 314 pp.

483 Boulanger, J. & McLellan, B. 2001: Closure violation in DNA-based capture-recapture
484 estimation of grizzly bear populations.- *Canadian Journal of Zoology* 79: 642-651.

485 Boulanger, J., White, G.C., McLellan, B., Woods, J., Proctor & M., Himmer, S. 2002: A
486 meta-analysis of grizzly bear DNA capture-recapture projects in British Columbia, Canada.-
487 *Ursus* 13: 137-152.

488 Boulanger, J., Himmer, S. & Swan, C. 2004: Monitoring of grizzly bear population trends and
489 demography using DNA capture-recapture methods in the Owikeno Lake area of British
490 Columbia.- *Canadian Journal of Zoology* 82: 1267-1277.

491 Boulanger, J., Stenhouse, G. & Munro, R. 2004a: Sources of heterogeneity bias when DNA
492 capture-recapture sampling methods are applied to grizzly bear (*Ursus arctos*) populations.-
493 *Journal of Mammalogy* 85: 618-624.

494 Boulanger, J., McLellan, B.N., Woods, J.G., Proctor, M.F. & Strobeck, C. 2004b: Sampling
495 design and bias in DNA- based capture-capture-recapture population and density estimates of
496 grizzly bears. *Journal of Wildlife Management* 68: 457-469.

497 Boulanger, J., Proctor, M., Himmer, S., Stenhouse, G., Paetkau, D. & Cranston, J. 2006: An
498 empirical test for DNA capture-recapture sampling strategies for grizzly bears.- *Ursus* 17:
499 149- 158.

500 Boulanger, J., Kendall, K.C., Stetz, J.C., Roon, D.A., Waits, L.P. & Paetkau, D. 2008:
501 Multiple data sources improve DNA-based capture-recapture estimates of grizzly bears.-
502 *Ecological Applications* 18: 577-589.

503 Broquet, T. & Petit, E. 2004: Quantifying genotyping errors in non-invasive population
504 genetics.- *Molecular Ecology* 13: 3601-3608.

505 Broquet, T., Ménard, N. & Petit, E. 2007: Non-invasive population genetics: a review of
506 sample source, diet, fragment length and microsatellite motif effects on amplification success
507 and genotyping error rates.- *Conservation Genetics* 8: 249-260.

508 Burnham, K.P. and Anderson, D.R. 1998: Model selection and multimodel inference – a
509 practical information-theoretic approach. Springer ISBN 0-387-95364-7.

510 Chao, A. 1987: Estimating the population size for capture-recapture data with unequal
511 catchability.- *Biometrics* 43: 783-791.

512 Chao, A.L. & Jeng, S.L. 1992: Estimating population size for capture-recapture data when
513 capture probabilities vary by time and individual animal.- *Biometrics* 48: 201-216.

514 Choquet, R., Lebreton, J.-D., Gimenez, O., Reboulet, A.-M., Pradel, R. 2009: U-CARE:
515 Utilities for performing goodness of fit tests and manipulating CAPture-REcapture data.-
516 *Ecography* 32: 1071-1074.

517 Coleman, K. & Wilson, D.S. 1998: Shyness and boldness in pumpkinseed sunfish: individual
518 differences are context-specific.- *Animal Behaviour* 56: 927-936.

519 Creel, S., Spong, G., Sands, J.L., Rotella, J., Zeigle, J., Joe, L., Murphy, K.M., Smith, D.
520 2003: Population size estimation in Yellowstone wolves with error-prone noninvasive
521 microsatellite genotypes.- *Molecular Ecology* 12, 2003-2009.

522 Crespin, L., Choquet, R., Lima, M., Merritt, J., Pradel, R. 2008: Is heterogeneity of
523 catchability in capture-recapture studies a mere sampling artefact or a biologically relevant
524 feature of the population? *Population Ecology* 50, 247-256.

525 Cubaynes, S., Pradel, R., Choquet, R., Duchamp, C., Gaillard, J.-M., Lebreton, J.-D.,
526 Marboutin, E., Miquel, C., Reboulet, A.-M., Poillot, C., Taberlet, P., Gimenez, O. 2010:
527 Importance of accounting for detection heterogeneity when estimating abundance: the case of
528 French wolves.- *Conservation Biology* 24, 621-626.

529 Dingemanse, N.J., Both, C., van Noordwijk, A.J., Rutten, A.L. & Drent, P. 2003: Natal
530 dispersal and personalities in great tits (*Parus major*). *Proceedings of the Royal Society*
531 (London) 270: 741-747.

532 Dreher, B.P., Winterstein, S.R., Scribner, K.T., Lukacs, P.M., Etter, D.R., Rosa, G.J.M.,
533 Lopez, V.A., Libants, S. & Filcek, K.B. 2007: Non-invasive estimation of black bear
534 abundance incorporating genotyping errors and harvested bear.- *Journal of Wildlife*
535 *Management* 71: 2684-2693.

536 Ebert, C., Huckschlag, D., Schulz, H.K. & Hohmann, U. (2009): Can hair traps sample wild
537 boar (*Sus scrofa*) randomly for the purpose of non-invasive population estimation?- *European*
538 *Journal of Wildlife Research*, DOI 10.1007/s10344-009-0351-7.

539 Eggert, L.S., Eggert, J.A. & Woodruff, D.S. 2003: Estimating population size for elusive
540 animals: the forest elephants of Kakum National Park, Ghana.- *Molecular Ecology* 12: 1389-
541 1402.

542 Fickel, J. & Hohmann, U. 2006: A methodological approach for non-invasive sampling for

543 population size estimates in wild boars (*Sus scrofa*).- European Journal of Wildlife Research
544 52: 28-33.

545 Foran, D.R., Minta, S.C. & Heinemeyer, K.S. 1997: DNA-based analysis of hair to identify
546 species and individuals for population research and monitoring.- Wildlife Society Bulletin
547 25: 840-847.

548 Frantz, A.C., Pope, L.C., Carpenter, P.J., Roper, T.J., Wilson, G.J. & Delahay, R.J. 2003:
549 Reliable microsatellite genotyping of the Eurasian badger (*Meles meles*) using faecal DNA.-
550 Molecular Ecology 12: 1649-1661.

551 Frantz, A.C., Schaul, M., Pope, L.C., Fack, F., Schley, L., Muller, C.P. & Roper, T.J. 2004:
552 Estimating population size by genotyping remotely plucked hair: the Eurasian badger.-
553 Journal of Applied Ecology 41: 985-995.

554 Gervasi, V., Ciucci, P., Boulanger, J., Posillico, M., Sulli, C., Focardi, S., Randi, E. &
555 Boitani, L. 2008: A preliminary estimate of the Apennine brown bear population size based
556 on hair-snag sampling and multiple data source capture-recapture Huggins models.- Ursus 19:
557 105-121.

558 Guschanski, K., Vigilant, L., McNeilage, A., Gray, M., Kagoda, E. & Robbins, M.M. 2009:
559 Counting elusive animals: Comparing field and genetic census of the entire mountain gorilla
560 population of Bwindi Impenetrable National Park, Uganda.- Animal Conservation 142: 290-
561 300.

562 Ihaka, R. & Gentleman, R. 1996: R: A language for data analysis and graphics.- Journal of
563 Computational and Graphical Statistics 5: 299-314.

564 Immell, D. & Anthony, R.G. 2006: Estimation of black bear abundance using a discrete DNA
565 sampling device.- Journal of Wildlife Management 72: 324-330.

566 Jacob, G., Debrunner, R., Gugerli, F., Schmid, B. & Bollmann, K. 2009: Field surveys of
567 capercaillie (*Tetrao urogallus*) in the Swiss Alps underestimated local abundance of the

568 species as revealed by genetic analyses of non-invasive samples.- Conservation Genetics DOI
569 10.1007/s10592-008-9794-8.

570 Knapp, S.H., Craig, B.A. & Waits, L.P. 2007: Incorporating genotyping error into non-
571 invasive DNA-based capture-recapture population estimates.- Journal of Wildlife
572 Management 73: 598-604.

573 Kohn, M.H., York, E.C., Kamradt, D.A., Haught, G., Sauvajot, R.M. & Wayne, R.K. 1999:
574 Estimating population size by genotyping faeces.- Proceedings of the Royal Society (London)
575 266: 657-663.

576 Link, W.A. 2003: Nonidentifiability of population size from capture-recapture data with
577 heterogeneous detection probabilities.- Biometrics 59: 1123-1130.

578 Link, W.A. 2004: Individual heterogeneity and identifiability in capture-recapture models.-
579 Animal Biodiversity and Conservation 27: 87-91.

580 Long, R.A., Donovan, T.M., Mackay, P., Zielinski, W.J. & Buzas, J.S. 2007: Comparing
581 faeces detection dogs, cameras, and hair snares for surveying carnivores.- Journal of Wildlife
582 Management 71: 2018-2025.

583 Lukacs, P.M. & Burnham, K.P. 2005: Review of capture-recapture methods applicable to
584 non-invasive genetic sampling.- Molecular Ecology 14: 3909-3919.

585 Maudet, C., Luikart, G., Dubray, D., von Hardenberg, A., Taberlet, P. 2004: Low genotyping
586 error rates in wild ungulate faeces sampled in winter.- Molecular Ecology Notes 4, 772-775.

587 McKelvey, K.S., Schwartz, M.K. 2004: Genetic errors associated with population estimation
588 using non-invasive molecular tagging: problems and new solutions.- Journal of Wildlife
589 Management 68: 439-448.

590 McLoughlin, P.D., Cluff, H.D., Gau, R.J., Mulders, R., Case, R. & Messier, F. 2003: Effect of
591 spatial differences in habitat on home ranges of grizzly bears.- Écoscience 10: 11-16.

592 Meijer, T., Norén, K., Hellström, P., Dalén, L. & Angerbjörn, A. 2008: Estimating population
593 parameters in a threatened arctic fox population using molecular tracking and traditional field
594 methods.- *Animal Conservation* 11: 330-338.

595 Menkens, G.E. & Anderson, S.H. 1988: Estimation of small-mammal population size.-
596 *Ecology* 69: 1952-1959.

597 Mettke-Hofmann, C., Ebert, C., Schmidt, T., Steiger, S. & Stieb, S., 2005. Personality traits in
598 resident and migratory warbler species.- *Behaviour* 142, 1357–1375.

599 Miquel, C., Bellemain, E., Poillot, C., Bessière, J., Durand, A. & Taberlet, P. 2006: Quality
600 indexes to assess the reliability of genotypes in studies using non-invasive sampling and
601 multiple-tube approach.- *Molecular Ecology Notes* 6: 985-988.

602 Miller, C.R., Joyce, P. & Waits, L.P. 2005: A new method for estimating the size of small
603 populations from genetic capture-recapture data.- *Molecular Ecology* 14: 1991-2005.

604 Mills, L.S., Citta, J.J., Lair, K.P., Schwartz, M.K. & Tallmon, D.A. 2000: Estimating animal
605 abundance using non-invasive DNA sampling: promise and pitfalls.- *Ecological Applications*
606 10: 283-294.

607 Minta, S. & Mangel, M. 1989: A simple population estimate based on simulation for capture-
608 recapture and capture-resight data.- *Ecology* 70: 1738-1751.

609 Mowat, G. & Paetkau, D. 2002: Estimating marten *Martes americana* population size using
610 hair capture and genetic tagging.- *Wildlife Biology* 8: 201-209.

611 Mowat, G. & Strobeck, C. 2000: Estimating population size of Grizzly bears using hair
612 capture, DNA profiling, and mark- recapture analysis.- *Journal of Wildlife Management* 64:
613 183-193.

614 Mowat, G., Heard, D.C., Seip, D.R., Poole, K.G., Stenhouse, G. & Paetkau, D.W. 2005:
615 Grizzly *Ursus arctos* and black bear *Ursus americanus* densities in the interior mountains of
616 North America.- *Wildlife Biology* 11: 31-48.

617 Obbard, M.E., Howe, E.J., Kyle, C.J. 2010: Empirical comparison of density estimators for
618 large carnivores.- *Journal of Applied Ecology* 47: 76-84.

619 Otis, D.L., Burnham, K.P., White, G.C. & Anderson, D.R. 1978: Statistical inference for
620 capture-recapture experiments.- *Wildlife Monographs* 62, 135 pp.

621 Paetkau, D. 2003: An empirical exploration of data quality in DNA- based population
622 inventories.- *Molecular Ecology* 12: 1375-1387.

623 Petit, E. & Valière, N. 2006: Estimating population size with non-invasive capture-recapture
624 data.- *Conservation Biology* 20: 1062-1073.

625 Piggott, M.P. & Taylor, A.C. 2003: Remote collection of animal DNA and its applications in
626 conservation management and understanding the population biology of rare and cryptic
627 species.- *Wildlife Research* 30: 1-13.

628 Pledger, S. & Efford, M. 1998: Correction of bias due to heterogeneous capture probability in
629 capture-recapture studies in open populations.- *Biometrics* 54: 888-898.

630 Pollock, K.H., Nichols, J.D., Brownie, C. & Hines, J.E. 1982: Statistical inference for
631 capture-recapture experiments.- *Wildlife Monographs* 107, 97 pp.

632 Poole, K.G., Mowat, G. & Fear, D.A. 2001: DNA-based population estimate for grizzly bears
633 *Ursus arctos* in northeastern British Columbia, Canada.- *Wildlife Biology* 7: 105-115.

634 Pradel, R. 2005: Multievent, an extension of multistate capture-recapture models to uncertain
635 states.- *Biometrics* 61, 442-447.

636 Prigioni, C., Remonti, L., Balestrieri, A., Sgrosso, S., Priore, G., Mucci, N. & Randi, E. 2006:
637 Estimation of European otter (*Lutra lutra*) population size by fecal DNA typing in southern
638 Italy.- *Journal of Mammalogy* 87: 855-858.

639 Prugh, L.R., Ritland, C.E., Arthur, S.M. & Krebs, C.J. 2005: Monitoring coyote population
640 dynamics by genotyping faeces.- *Molecular Ecology* 14: 1585-1596.

641 Puechmaille, S.J. & Petit, E. 2007: Empirical evaluation of non-invasive capture-capture-
642 recapture estimation of population size based on a single sampling session.- Journal of
643 Applied Ecology 44: 843-852.

644 Roon, D.A., Thomas, M.E., Kendall, K.C. & Waits, L.P. 2005a: Evaluating mixed samples as
645 a source of error in non-invasive genetic studies using microsatellites.- Molecular Ecology 14:
646 195-201.

647 Roon, D.A., Waits, L.P., Kendall, K.C. 2005b: A simulation test of the effectiveness of
648 several methods for error-checking non-invasive genetic data.- Animal Conservation 8, 203-
649 215.

650 Rudnick, J.A., Katzner, T.E., Bragin, E.A. & DeWoody, J.A. 2008: A non-invasive genetic
651 evaluation of population size, natal philopatry, and roosting behaviour of non-breeding
652 eastern imperial eagles (*Aquila heliaca*) in central Asia.- Conservation Genetics 9: 667-676.

653 Ruis, M.A.W., te Brake, J.H.A., van de Burgwal, J.A., de Jong, I.C., Blokhuis & H.J.,
654 Koolhaas, J. 2000: Personalities in female domesticated pigs: behavioural and physiological
655 indications.- Applied Animal Behavioural Sciences 66: 31-47.

656 Scheppers, T.L.J., Frantz, A.C., Schaul, M., Engel, E., Breyne, P., Schley, L. & Roper, T.J.
657 2007: Estimating social group size of Eurasian badgers *Meles meles* by genotyping remotely
658 plucked single hairs.- Wildlife Biology 13: 195-207.

659 Seber, G.A.F. 1982: The Estimation of Animal Abundance and Related Parameters. 2nd
660 edition, Charles Griffin, London, 654 pp.

661 Settlage, K.E., Van Manen, F.T., Clark, J.D. & King, T.L. 2008: Challenges of DNA-based
662 capture-recapture studies of American black bears.- Journal of Wildlife Management 72:
663 1035-1042.

664 Solberg, K.H., Bellemain, E., Drageset, O.-M., Taberlet, P. & Swenson, J.E. 2006: An
665 evaluation of field and non-invasive genetic methods to estimate brown bear (*Ursus arctos*)
666 population size.- Biological Conservation 128: 158-168.

667 Sweitzer, R., van Vuren, D., Gardner, I.A., Boyce, W. & Waithman, J.D. 2000: Estimating
668 sizes of wild pig Populations in the north and central coast regions of California.- Journal of
669 Wildlife Management 64: 531-543.

670 Taberlet, P., Waits, L.P. & Luikart, G. 1999: Non-invasive genetic sampling: look before
671 you leap.- Trends in Ecology and Evolution 14: 323-327.

672 Thompson, W.L., White, G.C. & Gowan, C. 1998: Monitoring Vertebrate Populations.-
673 Academic Press Inc., San Diego, 365 pp.

674 Triant, D.A., Pace, R.M. III. & Stine, M. 2004: Abundance, genetic diversity and
675 conservation of Louisiana black bears (*Ursus americanus luteolus*) as detected through non-
676 invasive sampling.- Conservation Genetics 5: 647-659.

677 Waits, L.P. & Paetkau, D. 2005: Noninvasive genetic sampling tools for wildlife biologists: a
678 review of applications and recommendations for accurate data collection.- Journal of Wildlife
679 Management 69: 1419-1433.

680 Wasser, S.K., Davenport, B., Ramage, E.R., Hunt, K.E., Parker, M., Clarke, C. & Stenhouse,
681 G. 2004: Faeces detection dogs in wildlife research and management: application to grizzly
682 and black bears in the Yellowhead Ecosystem, Alberta, Canada. Canadian Journal of Zoology
683 82: 475-492.

684 White, G.C., Andersen, D.R., Burnham, K.P. & Otis, D.L. 1982: Capture-recapture and
685 removal methods for sampling closed populations.- Los Alamos National Laboratory Report
686 LA-8787-NERP, Los Alamos, New Mexico, 235 pp.

687 White, G.C. & Burnham, K.P. 1999: Program MARK: survival estimation from populations
688 of marked animals.- Bird Study 46: 120-138.

689 Williams, B.K., Nichols, J.D. & Conroy, M.J. 2002: Analysis and management of animal
690 populations.- Academic Press, San Diego, California, USA, 817 pp.

691 Wilson, G.J., Frantz, A.C, Pope, L.C., Roper, T.J., Burke, T.A., Cheeseman, C.L. & Delahay,
692 R.J. 2003: Estimation of badger abundance using faecal DNA typing.- Journal of Applied
693 Ecology 40: 658-666.

694 Woods, J.G., Paetkau, D., Lewis, D., McLellan, B.N., Proctor, M. & Strobeck, C. 1999:
695 Genetic tagging of free-ranging black and brown bears.- Wildlife Society Bulletin 27: 616-
696 627.

697

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723 Table 1: Overview over all reviewed articles. For each article, every single population estimate is registered separately (e.g. in studies which were carried out over several years
724 or in different study areas). P average is the average per session capture probability reported in the studies (n.r. = not reported). For each study, it is noted if individual
725 heterogeneity (IH) was detected via goodness-of-fit (GOF) testing or model selection procedure.
726
727

Reference (study)	Species	Sampling	No. individuals sampled	p average	Estimated N	No. sampling occasions	Software used	IH detected in the field? (test method)	IH detected via GOF (test)
Banks et al. 2003	<i>Lasiorhinus kreftii</i>	hair	81	n.r.	113	7	CAPTURE	-	yes, CF (none)
Mowat & Paetkau 2002	<i>Martes americana</i>	hair	88	0.15	213	4	CAPTURE	-	yes (Chi ²)
Scheppers et al. 2007	<i>Meles meles</i>	hair	55	n.r.	61	1	CAPWIRE	yes (direct sightings)	yes (LR)
Triant et al. 2004 (coast sampling)	<i>Ursus americanus</i>	hair	57	n.r.	77	2	n.r.	-	no (-)
Triant et al. 2004 (inland sampling)		hair	32	n.r.	41	2	n.r.	-	no (-)
Boersen et al. 2003	<i>Ursus americanus</i>	hair	58	n.r.	119	14	CAPTURE	-	yes (Chi ²)
Immell & Anthony 2007 (Steamboat 2003)	<i>Ursus americanus</i>	hair	32	0,4	46	3	CAPTURE	-	yes (Chi ²)
Immell & Anthony 2007 (Steamboat 2004)		hair	30	0,2	57	3	CAPTURE	-	no (Chi ²)
Immell & Anthony 2007 (Toketee 2003)		hair	47	0,3	67	3	CAPTURE	-	yes (Chi ²)
Immell & Anthony 2007 (Toketee 2004)		hair	46	0,31	65	3	CAPTURE	-	yes (Chi ²)
Belant et al. 2005 (Stockton Island)	<i>Ursus americanus</i>	hair	26	0,68	26	4	CAPTURE	-	no (Chi ²)
Belant et al. 2005 (Sand Island)		hair	6	0,54	6	4	CAPTURE	-	sample size too small for test
Settlage et al. 2008 (National Park)	<i>Ursus americanus</i>	faeces	129	0,06 to 0,18	97 to 114	10	CAPTURE	-	yes (Chi ²)
Settlage et al. 2008 (National Forest)		faeces	60	0,09 to 0,32	197 to 330	10	CAPTURE	-	yes (Chi ²)
Poole et al. 2001 ⁽¹⁾	<i>Ursus arctos</i>	hair	98	0,19	148	5	CAPTURE	-	no (Chi ²)
Boulanger & McLellan 2001 ⁽¹⁾ (CAPTURE) ⁽²⁾	<i>Ursus arctos</i>	hair	98	n.r.	155	5	CAPTURE	-	no (Chi ²)

Boulanger & McLellan 2001 (MARK) ⁽¹⁾		hair	98	0,05 to 0,4	191	5	MARK	-	yes (AIC)
Boulanger et al. 2002 (Jumbo project)		hair	33	0,26	45	4	CAPTURE	-	yes (Chi ²)
Boulanger et al. 2002 (U. Columbia 97)		hair	40	0,2	55	5	CAPTURE	-	yes (Chi ²)
Boulanger et al. 2002 (U. Columbia 98)		hair	40	0,12	92	5	CAPTURE	-	yes (Chi ²)
Boulanger et al. 2002 (Kingcome)	<i>Ursus arctos</i>	hair	58	0,2	102	5	CAPTURE	-	yes (Chi ²)
Boulanger et al. 2002 (U. Columbia 96)		hair	55	0,16	108	4	CAPTURE	-	yes (Chi ²)
Boulanger et al. 2002 (Granby Kettle)		hair	22	0,13	46	5	CAPTURE	-	no (Chi ²)
Boulanger et al. 2002 (Prophet)		hair	98	0,17	166	5	CAPTURE	-	yes (Chi ²)
Boulanger et al. 2004a	<i>Ursus arctos</i>	hair	41	0,35	104	3	MARK	yes (radiotelemetry)	yes (AIC)
Boulanger et al. 2004b (sampling 1996)		hair	54	0,16	108	4	CAPTURE		no (AIC)
Boulanger et al. 2004b (sampling 1997)	<i>Ursus arctos</i>	hair	41	0,2	55	5	CAPTURE	yes (radiotelemetry)	no (AIC)
Boulanger et al. 2004b (sampling 1998)		hair	39	0,12	92	5	CAPTURE		no (AIC)
Bolanger et al. 2006	<i>Ursus arctos</i>	hair	41	0,32	43	4	MARK	-	yes (AIC)
Boulanger et al. 2004	<i>Ursus arctos</i>	hair	total: 123	n.a.	per year, mean 49	5	MARK	-	no (-)
Mowat & Strobeck 2000 (British Columbia)	<i>Ursus arctos</i>	hair	109	0,1	257	5	CAPTURE	-	no (Chi ²)
Mowat & Strobeck 2000 (Alberta)		hair	37	0,16	74	4	CAPTURE	-	yes (Chi ²)
Mowat et al. 2005 (sampling SC Selkirks)		hair	38	0,09	97	5	CAPTURE	-	?
Mowat et al. 2005 (sampling NC Selkirks)		hair	74	0,08	223	5	CAPTURE	-	?
Mowat et al. 2005 (sampling Prophet plateau)		hair	32	0,13	63	5	CAPTURE	-	no (Chi ²)
Mowat et al. 2005 (sampling Prophet Mtns)	<i>Ursus arctos</i>	hair	67	0,21	96	5	CAPTURE	-	no (Chi ²)
Mowat et al. 2005 (sampling Yellowhead)		hair	48	0,16	107	4	CAPTURE	-	?
Mowat et al. 2005 (sampling Parsnip Plateau)		hair	21	0,12	50	4	CAPTURE	-	no (Chi ²)
Mowat et al. 2005 (sampling Parsnip Mtns)		hair	216	0,22	341	4	CAPTURE	-	no (Chi ²)

Mowat et al. 2005 (sampling Parsnip Plateau)	<i>Ursus americanus</i>	hair	194	0,06	892	4	CAPTURE	-	?
Mowat et al. 2005 (sampling Parsnip Mtns)		hair	85	0,08	363	4	CAPTURE	-	?
Mowat et al. 2005 (sampling Bowron river)	<i>Ursus arctos</i>	hair	53	0,32	76	3	CAPTURE	-	?
Woods et al. 1999	<i>Ursus arctos, U. americanus</i>	hair	54	0.05	104	4	CAPTURE	-	no (-)
Gervasi et al. 2008 ⁽³⁾	<i>Ursus arctos marsicanus</i>	hair	30	0,03	44	12	MARK	-	no (-)
Frantz et al. 2004	<i>Meles meles</i>	hair	14	n.r.	12 to 19	n.r.	GIMLET/ R (rarefaction)	no (direct counts)	no (-)
Dreher et al. 2007	<i>Ursus americanus</i>	hair + harvested bears	544	0,02 (hair), 0,21 (hunt)	1882	5	MARK	radiotelemetry, harvested bears (hair: yes)	no (-)
Wasser et al. 2004	<i>Ursus arctos, U. americanus</i>	hair, faeces	24	n.r.	28	5	MARK	radiotelemetry, comparison between sampling methods (no)	no (-)
Cubaynes et al. 2010	<i>Canis lupus</i>	faeces	160	0.01 to 0.86	3 to 126	n.r.	E-SURGE, U-CARE	-	yes (AIC)
Prugh et al. 2005	<i>Canis latrans</i>	faeces	total: 56	0,6 and 0,75	per year, mean 26	9	MARK	radiotelemetry (yes)	yes (AIC, CF)
Prigioni et al. 2006	<i>Lutra lutra</i>	faeces	23	n.r.	36	n.r.	SPSS (rarefaction)	-	yes, CF (none)
Puechmaille & Petit 2007 (sampling Epiniac 2003)	<i>Rhinolophus hipposideros</i>	faeces	54	n.r.	approx. 65	1	R	direct counts (no)	no (LR, S)
Puechmaille & Petit 2007 (sampling Pluherlin 2003)		faeces	35	n.r.	approx. 42	1	R		no (LR, S)
Puechmaille & Petit 2007 (sampling Saint-Thurial 2003)		faeces	16	n.r.	approx. 28	1	R		yes (LR, S)
Puechmaille & Petit 2007 (sampling Epiniac 2004)		faeces	58	n.r.	approx. 85	1	R		yes (LR, S)
Puechmaille & Petit 2007 (sampling Pluherlin 2004)		faeces	35	n.r.	approx. 48	1	R		yes (LR, S)
Puechmaille & Petit 2007 (sampling Saint-Thurial 2004)		faeces	14	n.r.	approx. 18	1	R		no (LR, S)
Frantz et al. 2003 ⁽⁴⁾		<i>Meles meles</i>	faeces	20	0,15	26	10		CAPTURE
Bellemain et al. 2005 (sampling 2001)	<i>Ursus arctos</i>	faeces	311	n.r.	approx. 480	11	MARK	-	yes (AIC)
Bellemain et al. 2005 (sampling 2002)		faeces	239	n.r.	approx. 350	13	MARK	-	yes (AIC)
Bellemain et al. 2007	<i>Ursus arctos</i>	faeces	28	n.r.	32 and 47	n.r.	GIMLET/ R (rarefaction)	-	no (-)

Solberg et al. 2006 (sampling 2001)	<i>Ursus arctos</i>	faeces	146	n.r.	223	11	MARK	-	yes (AIC)	
Solberg et al. 2006 (sampling 2002)		faeces	81	n.r.	157	13	MARK	-	yes (AIC)	
Jacob et al. 2009 (sampling Obwalden)	<i>Tetrao urogallus</i>	faeces	29	n.r.	78	1	CAPWIRE	-	yes (LR)	
Jacob et al. 2009 (sampling Regelstein)		faeces	9	n.r.	14	1	CAPWIRE	-	yes (LR)	
Jacob et al. 2009 (sampling Höhi)		faeces	16	n.r.	20	1	CAPWIRE	-	no (LR)	
Jacob et al. 2009 (sampling Schwägälp)		faeces	7	n.r.	10	1	CAPWIRE	-	yes (LR)	
Jacob et al. 2009 (sampling Rofla)		faeces	7	n.r.	10	1	CAPWIRE	-	yes (LR)	
Jacob et al. 2009 (sampling Salouf)		<i>Tetrao urogallus</i>	faeces	5	n.r.	5	1	CAPWIRE	-	no (LR)
Jacob et al. 2009 (sampling Albula-north)		faeces	23	n.r.	36	1	CAPWIRE	-	no (LR)	
Jacob et al. 2009 (sampling Albula-south)		faeces	8	n.r.	33	1	CAPWIRE	-	no (LR)	
Eggert et al. 2003	<i>Loxodonta cyclotis</i>	faeces	86	0,0 to 0,2	225	10	CAPTURE	-	yes (Chi ²)	
Banks et al. 2002	<i>Vombatus ursinus</i>	faeces	17	n.r.	19	5	CAPTURE	-	yes (Chi ²)	
Rudnick et al. 2008	<i>Aquila heliaca</i>	faeces, feathers	278	> 0,44	308	4	MARK	-	yes, CF (none)	
Marucco et al. 2009 ⁽⁵⁾	<i>Canis lupus</i>	faeces	total: 87	0,28 to 0,77	per year, 21 to 47	14	MSURGE, UCARE	yes ⁽⁵⁾ (snow tracking, survey of marking behaviour)	no (AIC)	
Meijer et al. 2008	<i>Alopex lagopus</i>	faeces	30	n.r.	42	2	n.r.	-	no (-)	
Wilson et al. 2003 ⁽⁴⁾	<i>Meles meles</i>	faeces	20	n.r.	36	4	Proc NLIN	video control, marked badgers (yes)	yes, CF (none)	
Kohn et al. 1999	<i>Canis latrans</i>	faeces	30	n.r.	41	n.r.	JMP IN3	radiotelemetry (yes)	yes, CF (none)	

LR = Likelihood ratio test (see Miller et al. 2005) implemented in model selection procedure of program CAPWIRE

S = simulation test

Chi² = Chi²- tests implemented in Program CAPTURE model selection routine (Otis et al. 1978)

AIC = IH suggested by the model selection process (via Akaike's Information Criterion) in program MARK (White & Burnham 1999)

CF = Capture frequency, i.e. IH is visible in the number of times different individuals are sampled

728

729

730

731

732

733

734
735
736
737
738
739
740
741

⁽¹⁾ Both articles deal with the same grizzly bear data set (Prophet River Project 1998). In both, Chi²- tests failed to detect IH. But Boulanger & McLellan (2001) detected IH as a function of distance from study area (via AIC) as well as assumed age-specific capture probabilities.

⁽²⁾ For Boulanger & McLellan 2001, two different population estimation methods are reported for the same data set. Because of their differing results, both are listed here separately

⁽³⁾ In this paper, heterogeneity was modelled for each of three different hair sampling approaches applied, but not IH as discussed in our review due to small sample size. Thus, IH was assumed by the authors to be present in the data set, but not detected via tests.

⁽⁴⁾ Both articles deal with the same badger data set. In Frantz et al. 2003, results of program CAPTURE's Chi²- tests are not reported, even though the authors note that they assume the heterogeneity models to be most appropriate. Wilson et al. reported the capture frequencies of individual badgers which revealed IH.

⁽⁵⁾ IH was detected, but it did not impact the population estimate – for details see text

Table 2: Support of logistic regression models testing the impact of coverage (i.e. the ratio on the detection of individual heterogeneity in the reviewed studies ('cov' = coverage, 'cov2' = squared coverage, 'samp_occ' = number of sampling occasions))

Model	AIC		Δ AIC	
	weighted		unweighted	
p, p2	27,49	0	14,30	0
p, p2, samp_occ	29,40	1,90	16,12	1,82
p, p2, samp_occ, p* samp_occ, p2* samp_occ	32,64	5,15	19,80	5,50
Null	36,00	8,51	24,08	9,78
samp_occ	37,45	9,95	25,81	11,51
P	37,45	9,96	24,41	10,11

Table 3: Support of logistic regression models testing the impact of capture probability on the detection of individual heterogeneity in the reviewed studies ('p' = capture probability, 'p2' = squared capture probability, 'samp_occ' = number of sampling occasions).

Model	AIC	Δ AIC
p, p2	27,49	0
p, p2, samp_occ	29,40	1,90
p, p2, samp_occ, p*samp_occ, p2*samp_occ	32,64	5,15
Null	36,00	8,51
samp_occ	37,45	9,95
P	37,45	9,96

Table 4: Results of the regression model testing for the impact of capture probability on the detectability of individual heterogeneity with best fit according to AIC ranking (model 'p, p2').

Parameter	estimate	SE	p
intercept	-12.071	5.856	0.0393
p	96.611	49.707	0.0519
p2	-161.981	89.441	0.0701

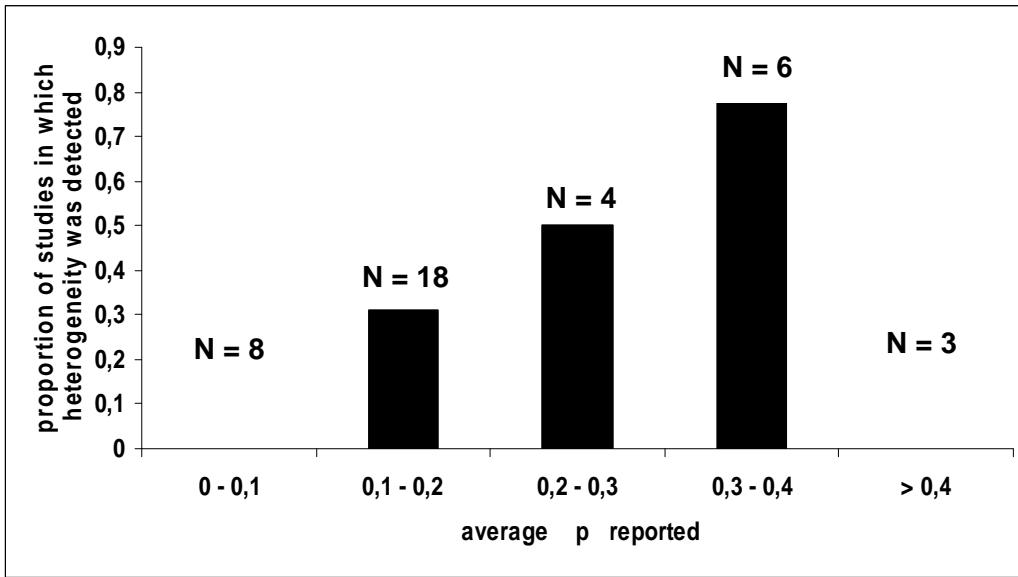


Fig. 1: Proportion of population estimates in which individual sampling heterogeneity was detected via goodness-of-fit test in relation to average capture probability (average p reported). Only those studies are included which report an estimate of the average capture probability p (39 of 75 studies, see table 1).

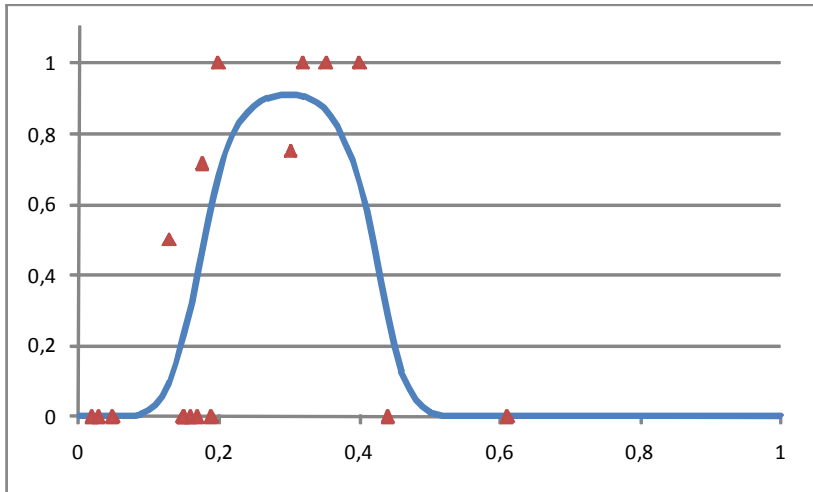


Fig 2: Observed (triangles) and predicted (line) values of IH detectability dependent on p and p2.