

1 **Breast cancer screening in women taking hormone replacement therapy** 2 **needs updating.**

3 Running title: Breast Cancer Screening and HRT

4

5 Philippe R. Koninckx ¹, Anastasia Ussia ², Geert Page ³

6 ¹ *Prof emeritus Obstetrics and Gynecology KULeuven, Leuven, Belgium, the*

7 *University of Oxford, Oxford, UK, University Cattolica, Rome, Italy and Moscow*

8 *State University, Moscow, Russia.* ² *Gruppo Ialo Belga, Villa del Rosario, Rome*

9 *Italy.* ³ *OBGYN, MSc EBHC, Coordinator Clinical Guidance project VVOG, Belgium.*

10

11 **Correspondence at:** Philippe R. Koninckx, Vuilenbos 2, 3360 Bierbeek, Belgium.

12 E-mail: Pkoinckx@gmail.com

13 **Abstract**

14 Breast cancer screening by mammography is widely used. The diagnostic
15 accuracy is limited, with a positive predictive value of 16%. Therefore, a stepwise
16 investigation, with repeat mammography and confirmation by pathology, is
17 usually proposed. Although this stepwise investigation intends to avoid
18 overtreatment, the many false positives result in unnecessary fear and
19 diagnostic surgery in many women. The false negatives are not known since
20 these women have not been investigated. Given the estimated low risk of
21 missing breast cancer and the slow growth, repeating a screening
22 mammography every two years is sufficient.

23 The false positive screening results increase with breast density, and breast
24 density increases when hormone replacement therapy (HRT) is given. It,
25 therefore, is suggested to use clinical judgment and stop HRT for 3 to 6 months
26 before repeating the mammography instead of starting immediately a stepwise
27 investigation in all women.

28

29 **Introduction**

30 Mammography is widely used for breast cancer screening. The risk of
31 overdiagnosis, overtreatment, and the many factors involved (Ryser et al., 2022),
32 and the benefits measured by the estimated lifetime gained (Bretthauer et al.,
33 2023) have been widely discussed. Less emphasised is that the reported
34 sensitivities and specificities, varying between 73% and 88% (Schünemann et
35 al., 2020) to 80% and 98% (Mushlin et al., 1998), result, at best, in positive
36 predictive values of less than 5% to 18% for prevalences of 0.5% in the
37 screened population (Fig 1). This is also illustrated by the Belgian breast cancer
38 screening program with biannual mammography in women between 50 and 69
39 years old, following The European quality assurance program (Schünemann et
40 al., 2020). Data from 2017 to 2020 illustrate that 5% of women screened, had a
41 second mammography, and 2% had more exams, such as MRI or biopsies, to
42 detect between 0.5% and 0.6% breast cancers. Thus, out of 200 women
43 screened, 10 are selected for a second mammography and 4 for more invasive
44 exams to find 1 cancer (Belgium, 2021).

45 This stepwise procedure compensates for the low predictive values of
46 mammography with sensitivities and specificities of 67% and 98% when
47 prevalences are low (Koninckx et al., 2023). Sensitivities and specificities are test
48 characteristics, while clinically, it is important to know the risk of having breast
49 cancer if the test is positive and the risk of missing cancer when the test is
50 negative. These are the predictive values, which decrease sharply when
51 prevalences are less than 10%, the PPV being the $\frac{\text{sensitivity} \times \text{prevalence}}{\text{sensitivity} \times \text{prevalence} + (1 - \text{specificity})(1 - \text{prevalence})}$ (Lesaffre and Lawson,
52 2012). Therefore, excellent sensitivities and specificities around 80% and 98%
53 result in poor predictive values when prevalences are 0.5%.

54 Although the interpretation of mammographies and the definition of abnormal
55 findings or suspicious lesions seem to be established (Gøtzsche and Jørgensen,

572013), there is little consensus about screening ages and intervals between
58screening (Schünemann et al., 2020). Without discussing ductal carcinoma in
59situ, the clinical benefits of screening or contrast mammography, we must
60realise the difficulties of performing randomised controlled trials for breast
61cancer screening and of translating the findings into guidelines. False and true
62negatives are poorly known since these women are not investigated. Evidence-
63based medicine and guidelines rely heavily on double-blind randomised
64controlled trials. However, RCTs are poorly suited to investigate multivariate
65events because of randomisation problems. A 2 or 3 Y/N factorial design already
66requires 4 or 8 groups. An RCT is not suited for rare events, and breast cancer
67trials require huge numbers taking time to perform with the risk of being
68outdated before being finished. This explains the lack of data on newer
69techniques in ultrasound, digital imaging, deep learning (Arun Kumar and
70Sasikala, 2023) and artificial intelligence (Gao et al., 2023). The translation of
71predictive values or Bayesian probabilities, into guidelines, results from
72estimating truth or clinical importance, decided by consensus or voting by a
73group of experts (Kubota et al., 2023), thus introducing subjectivity based on
74previous experiences. Therefore, guidelines may vary, although based on the
75same RCTs (Jhangiani et al., 2023). More fundamental is that accuracies and
76predictive values of abnormal or suspicious findings are crude estimates and
77should be stratified for factors such as breast densities, increasing the difficulty
78of interpretation and the risk of abnormal findings (Freer, 2015, Schünemann et
79al., 2020). Finally, although beyond this discussion, it should be realised that the
80accuracy of the diagnosis by pathology cannot be established since it is the gold
81standard, which cannot be compared with a better test. In addition, it is unclear
82whether histopathological parameters reflect the expression profile or molecular
83genetic features (Simpson et al., 2005).

84 These poor predictive values of mammography and the stepwise clinical
85 approach result in more lumpectomies and mastectomies (Gøtzsche and
86 Jørgensen, 2013, Schünemann et al., 2020) with an estimated 30% overdiagnosis
87 out of prudence. Without discussing whether screening results in a reduction in
88 breast cancer mortality, we should realise that a reduction of breast cancer
89 mortality by 15% and an overdiagnosis and overtreatment of 30% results in
90 preventing one mortality for every 2000 women screened for ten years, but at
91 the cost of 10 women being treated unnecessarily, and 200 women having
92 experienced psychological distress and anxiety. Although the exact figures can
93 be discussed, the conclusion of overdiagnosis and little effect on mortality seems
94 repetitively confirmed (Bretthauer et al., 2023)

95 **Breast cancer screening, clinical medicine and Bayesian statistics**

96 Clinical medicine is multivariate, and in the individual woman (Koninckx et al.,
97 2023), the probability of having breast cancer varies not only with the image of
98 mammography (or MRI) but also with the clinical exam, age, heredity, obesity,
99 breast density, race and many other factors. The estimation of the probability in
100 the individual woman that an abnormal or suspicious lesion is a cancer should
101 include all these factors besides the mammography. This requires multivariate
102 Bayesian statistics, but these results are unfortunately not available. For
103 example, since multivariate RCTs are difficult to perform, the added value of
104 mammography following a negative clinical exam, with a reported sensitivity of
105 54% and specificity of 94% (Jatoi, 2003), is still unknown. Therefore, the
106 translation of an abnormal or suspicious lesion into the management of the
107 individual woman remains a clinical judgment. This explains that
108 recommendations for the diagnostic workup of abnormal or suspicious breast
109 lesions vary.

110 **Breast cancer screening and hormone replacement therapy**

111Hormone replacement therapy (HRT) increases breast density and thus the risk
112of finding abnormal and suspicious lesions (Rutter et al., 2001, Freer, 2015, Azam
113et al., 2018). Therefore, it was suggested that HRT be stopped for a few months
114before screening mammography is performed (Beckmann et al., 2013).

115Unfortunately stopping HRT for several months is clinically poorly accepted.
116Since breast cancer screening by mammography is recommended to be
117performed only every two years, the risk of delaying the workup and eventual
118treatment of abnormal or suspicious lesions for 3 to 6 months must be very low.
119Therefore, managing women on HRT and abnormal or suspicious lesions on
120screening mammography must be clinically individualised. Considering all risk
121factors, the increase of abnormal or suspicious images in women taking HRT and
122the low risk of delaying therapy for a few months, it may be considered to stop
123HRT for 3 to 6 months before repeating mammography. The women whose
124repeat mammography returns to normal will have avoided unnecessary cancer
125workups and surgery, and the continuation of HRT can be discussed.

126**Conclusion and discussion**

127The limitations of performing multivariate RCTs and the frequentist statistical
128analysis hamper the interpretation of breast cancer screening by mammography.
129It is still not fully appreciated that the predictive value of any test is strongly
130affected by the prevalence of the disease (Lesaffre and Lawson, 2012). For
131breast cancer with a prevalence of 0.5%, the positive predictive value is only
13216% for sensitivities of 67% and specificities of 98%. Unfortunately, negative
133predictive values cannot be estimated since the women in whom breast cancer is
134missed (false negatives) are unknown since they have not been further explored.
135Also, although estimated as high as 40% (Ryser et al., 2022), the true incidence
136of false positives is not clear since expectant management of suspicious lesions
137is considered unethical. Clinically important is that frequentist statistical analysis

138calculates the positive predictive values of breast cancer screening but not the
139added value besides age, heredity and breast density. This would require a
140Bayesian statistical approach, which has not been performed yet to the best of
141our knowledge.

142The diagnostic workup of abnormal and suspicious lesions follows a clinical logic
143of repeat exams and, eventually, biopsies, with the dogma (without data) that
144delaying the workup and treatment can only be harmful. Overtreatment and
145unnecessary fear have been discussed, but the hypothesis that some lesions
146could disappear spontaneously, especially in women stopping HRT, has not been
147tested.

148It is beyond this manuscript to discuss the limitations of frequentist statistics or
149the need to individualise therapy and to consider all predictive factors or medical
150and corporate pressures, as we recently did for endometriosis (Koninckx et al.,
1512022). However, individualising follow-up and postponing exams for three
152months in women with abnormal or suspicious lesions and dense breasts while
153HRT is stopped has the potential to decrease overtreatment. Although it would
154be an easy trial to demonstrate that some abnormal images become normal
155after stopping HRT for three months, it is close to impossible to demonstrate that
156a delay of 3 months does not increase the risk since proving the absence of an
157effect requires huge numbers. Therefore, an adequately performed RCT to
158demonstrate a decreased risk of unnecessary treatments without increasing risks
159is unlikely to be performed. In conclusion, individualising follow-up and letting
160the woman decide whether to stop HRT for three or more months before
161repeating mammography has the potential to decrease overtreatment.

162

163Word count: 1196

164

165 **Disclosure:** None of the authors has a conflict of interest to declare.

166 **IRB board:** The IRB board of KU Leuven Gasthuisberg confirmed in writing that
167 for experiments not involving human or animal tissues, IRB board approval was
168 not requested

169 **Authors role.** All authors discussed the manuscript and agreed with the final
170 manuscript.

171 **Funding:** No funding

172

173 **List of legends**

174 Fig 1 Positive predictive values ($PPV = \frac{\text{sensitivity} \times \text{prevalence}}{\text{sensitivity} \times \text{prevalence} + (1 - \text{specificity})(1 - \text{prevalence})}$) of a mammography if

175 sensitivities and specificities are 73% and 88% or 80% and 98%. This

176 illustrates the crucial importance of the specificity and of the prevalence of the

177 disease (0.5 and 1% indicated).

178

179

180

Reference List

181

182 Arun Kumar S and Sasikala S. Review on Deep Learning-Based CAD Systems for
183 Breast Cancer Diagnosis. Technol. Cancer Res. Treat.
184 2023;22:15330338231177977.

185

186 Azam S, Lange T, Huynh S, et al. Hormone replacement therapy, mammographic
187 density, and breast cancer risk: a cohort study. Cancer Causes & Control.
188 2018;29:495-505.

189

190 Beckmann KR, Farshid G, Roder DM, et al. Impact of hormone replacement
191 therapy use on mammographic screening outcomes. Cancer Causes & Control.
192 2013;24:1417-26.

193

194 Belgium (2021). "Jaarrapport bevolkingsonderzoeken naar kanker 2021."
195 Retrieved 03-09-2023, from <https://www.zorg-en-gezondheid.be/nieuws/jaarrapport-bevolkingsonderzoeken-naar-kanker-2021>.

197

198Bretthauer M, Wieszczy P, Løberg M, et al. Estimated Lifetime Gained With
199Cancer Screening Tests: A Meta-Analysis of Randomized Clinical Trials. JAMA
200Intern Med. 2023;

201

202Freer PE. Mammographic breast density: impact on breast cancer risk and
203implications for screening. Radiographics. 2015;35:302-15.

204

205Gao Y, Lin J, Zhou Y and Lin R. The application of traditional machine learning and
206deep learning techniques in mammography: a review. Front Oncol.
2072023;13:1213045.

208

209Gøtzsche PC and Jørgensen KJ. Screening for breast cancer with mammography.
210Cochrane Database of Systematic Reviews. 2013;

211

212Jatoi I. Screening clinical breast examination. The Surgical clinics of North
213America. 2003;83:789-801.

214

215Jhangiani N, Philip M and Jatoi I. Breast cancer screening guidelines:
216discrepancies raise concerns about validity. Breast Cancer. 2023;30:869-71.

217

218Koninckx PR, Ussia A, Alsuwaidi S, et al. Reconsidering evidence-based
219management of endometriosis. Facts Views Vis Obgyn. 2022;14:225-33.

220

221Koninckx PR, Wattiez A and Saridogan E. Statistics, truth finding and predictions:
222what every gynaecologist and researcher should know. Facts Views Vis Obgyn.
2232023;15:95-7.

224

225Kubota K, Nakashima K, Nakashima K, et al. The Japanese breast cancer society
226clinical practice guidelines for breast cancer screening and diagnosis, 2022
227edition. Breast Cancer. 2023;

228

229Lesaffre E and Lawson AB. Bayesian biostatistics. United Kingdom, John Wiley &
230Sons, ltd.2012.

231

232Mushlin AI, Kouides RW and Shapiro DE. Estimating the accuracy of screening
233mammography: a meta-analysis. Am. J. Prev. Med. 1998;14:143-53.

234

235Rutter CM, Mandelson MT, Laya MB, et al. Changes in breast density associated
236with initiation, discontinuation, and continuing use of hormone replacement
237therapy. JAMA. 2001;285:171-6.

238

239Ryser MD, Lange J, Inoue LYT, et al. Estimation of Breast Cancer Overdiagnosis in
240a U.S. Breast Screening Cohort. Ann Intern Med. 2022;175:471-8.

241

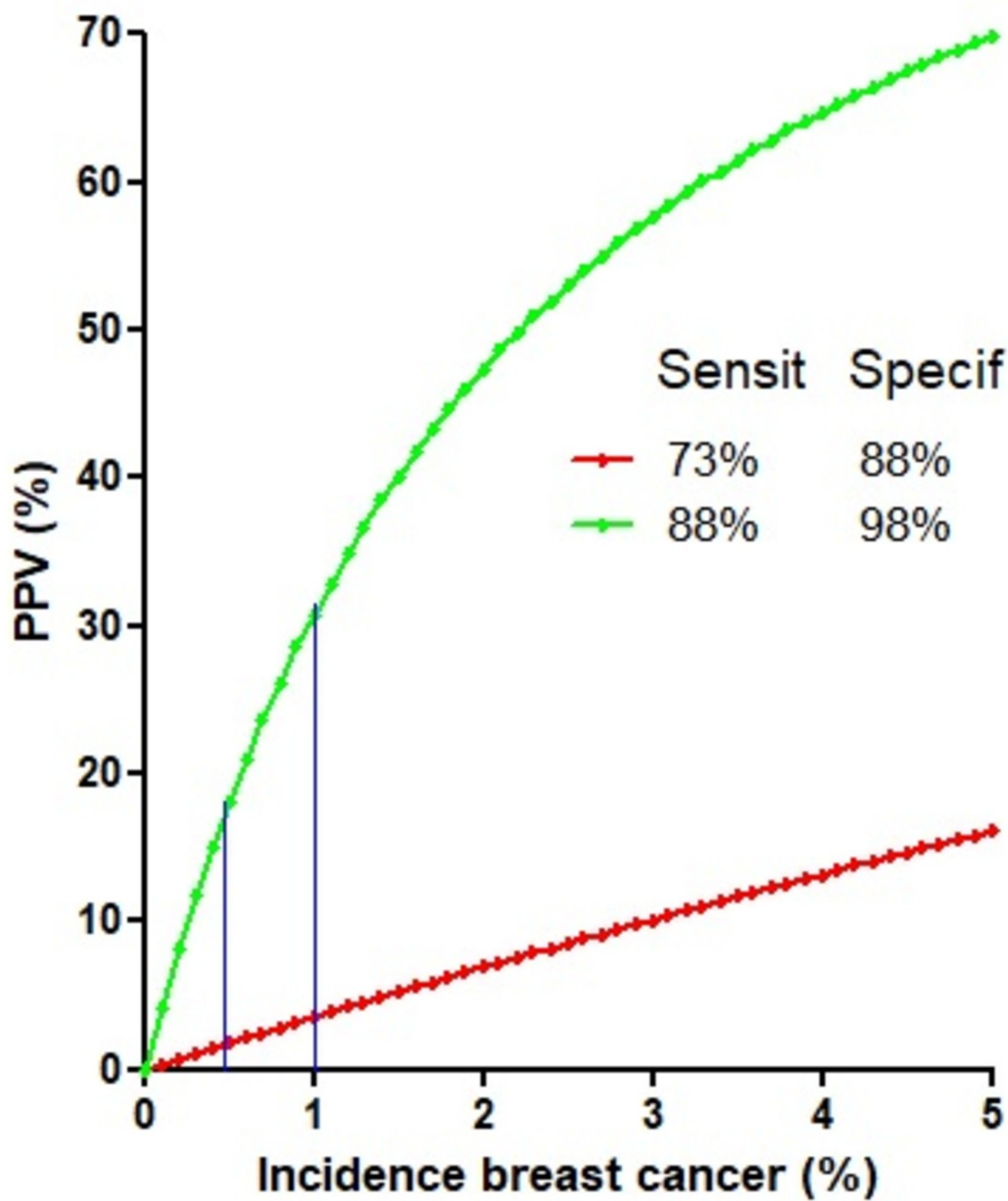
242Schünemann HJ, Lerda D, Quinn C, et al. Breast Cancer Screening and Diagnosis:
243A Synopsis of the European Breast Guidelines. Ann Intern Med. 2020;172:46-56.

244

245Simpson PT, Reis-Filho JS, Gale T and Lakhani SR. Molecular evolution of breast
246cancer. *The Journal of Pathology*. 2005;205:248-54.

247

248



1 Breast cancer screening in women taking hormone replacement therapy 2 needs updating.

3 Running title: Breast Cancer Screening and HRT

4

5 Philippe R. Koninckx ¹, Anastasia Ussia ², Geert Page ³

6 ¹ Prof emeritus Obstetrics and Gynecology KU Leuven, Leuven, Belgium, the

7 University of Oxford, Oxford, UK, University Cattolica, Rome, Italy and Moscow

8 State University, Moscow, Russia. ² Gruppo Ialo Belga, Villa del Rosario, Rome

9 Italy. ³ OBGYN, MSc EBHC, Coordinator Clinical Guidance project VVOG, Belgium.

10

11 **Correspondence at:** Philippe R. Koninckx, Vuilenbos 2, 3360 Bierbeek, Belgium.

12 E-mail: Pkoinckx@gmail.com

13 Abstract

14 Breast cancer screening by mammography is widely used. The diagnostic

15 accuracy is limited, with ~~only~~ a positive predictive value of 16%. Therefore, a

16 stepwise investigation, with repeat mammography and confirmation by

17 pathology, is usually proposed. Although this stepwise investigation intends to

18 avoid overtreatment, the many false positives result in unnecessary fear and

19 diagnostic surgery in many women. - The false negatives are not known since

20 these women have not been investigated. ~~Given However, considering~~ the

21 estimated low risk of missing breast cancer and the slow growth, repeating a

22 screening mammography every two years is ~~considered~~ sufficient.

23 The false positive screening results increase with breast density, and breast

24 density increases when hormone replacement therapy (HRT) is given. It,

25 therefore, is suggested to use clinical judgment ~~and to~~ stop HRT for 3 to 6 months

26 before repeating the mammography instead of starting immediately a stepwise

27 investigation in all women.

28

29 Introduction

30 Mammography is widely used for breast cancer screening, and the risk of
31 overdiagnosis, and overtreatment, and the many factors involved (Ryser et al.,
32 2022), and the benefits measured by the estimated lifetime gained (Bretthauer
33 et al., 2023) have been widely discussed. Less ~~emphasised~~ emphasised is that
34 the reported sensitivities and specificities, varying between 73% and 88%
35 (~~Abdullah et al., 2021~~) (Schünemann et al., 2020) to 80% and 98% (Mushlin et
36 al., 1998), resulting, at best, in ~~less than 5% to 18%~~ positive predictive values
37 of less than 5% to 18% for prevalences of 0.5% in the screened population (Fig
38 1). This is also illustrated by the Belgian breast cancer screening program with
39 biannual mammography in women between 50 and 69 years old, following The
40 European quality assurance program (Schünemann et al., 2020) (Schünemann et
41 al., 2020b) (Von Karsa and Arrossi, 2013) (Von Karsa and Arrossi, 2013) Data -
42 from 2017 to 2020 illustrate that in women between 50 and 69 years old, with
43 5% of women screened, had undergoing a second mammography, evaluation
44 and 2% had more exams, such as MRI or biopsies, to detect find between 0.5%
45 and 0.6% breast cancers. The Belgian stepwise screening program, Thus, out of
46 200 women screened, 10 are selected for a second mammography and 4 for
47 more invasive exams to find 1 cancer (Belgium, 2021).

48 This stepwise procedure compensates for the low predictive values of
49 mammography with sensitivities and specificities of 67 % and 98% when
50 prevalences are low (Koninckx et al., 2023) with a sensitivity of 67% and a
51 specificity of 98% (Belgium, 2021); thus has at best (prevalence 0.6%) a positive
52 predictive value of 16% and 29%, following a first or second positive screening,
53 respectively. Sensitivities and specificities are test characteristics, while
54 clinically, it is important to know the risk of having breast cancer if the test is
55 positive and the risk of missing cancer when the test is negative. These are the
56 predictive values, which decrease sharply when prevalences are less than 10%,

57 the PPV being the sensitivity*prevalence / (sensitivity*prevalence + (1-
58 specificity)(1-prevalence)) (Lesaffre and Lawson, 2012). Therefore, excellent
59 sensitivities and specificities around 80% and 98% result in poor predictive
60 values when prevalences are 0.5%.

61 Although the interpretation of mammographies and the definition of abnormal
62 findings or suspicious lesions seems to be established (Gøtzsche and Jørgensen,
63 2013), there is little consensus about screening ages and intervals between
64 screening (Schünemann et al., 2020). Without discussing ductal carcinoma in
65 situ, the clinical benefits of screening (Gøtzsche, 2015, Bretthauer et al., 2023) or
66 ~~the recent advances in reading mammography using artificial intelligence (Gao-~~
67 ~~et al., 2023)~~ or contrast mammography, we must ~~realise~~ realise the difficulties of
68 performing randomised controlled trials for breast cancer screening and of
69 uncertainty of translating the findings mammography as a screening test into-
70 sensitivity and specificity into guidelines. First, fFalse and true negatives are
71 poorly known since these women are not investigated. Evidence-based medicine
72 and guidelines rely heavily on double-blind randomised controlled trials.
73 However, RCTs are poorly suited to investigate multivariate events because of
74 randomisation problems. A 2 or 3 Y/N factorial design already requires 4 or 8
75 groups. An RCT is not suited for rare events, and breast cancer trials require
76 huge numbers taking time to perform with the risk of being outdated before
77 being finished. This explains the lack of data on newer techniques in ultrasound,
78 digital imaging, deep learning (Arun Kumar and Sasikala, 2023) and artificial
79 intelligence (Gao et al., 2023). The translation of predictive values or Bayesian
80 probabilities, into guidelines, results from estimating truth or clinical importance,
81 decided by consensus or voting by a group of experts (Kubota et al., 2023), thus
82 introducing subjectivity based on previous experiences. Therefore, guidelines
83 may vary, although based on the same RCTs (Jhangiani et al., 2023). Secondly,

84sensitivities and specificities are test characteristics, while clinically, it is
85important to know the risk of having breast cancer if the test is positive and the
86risk of missing cancer when the test is negative. These are the predictive values,
87which decrease sharply when prevalences are less than 10%, the PPV being the
88 $\text{sensitivity} \times \text{prevalence} / (\text{sensitivity} \times \text{prevalence} + (1 - \text{specificity})(1 - \text{prevalence}))$
89(Lesaffre and Lawson, 2012). Therefore, excellent sensitivities and specificities
90around 80% and 98% result in a poor predictive value of 16% when prevalences
91are 0.5%. This explains the sequential approach of retesting abnormal or
92suspicious lesions once or twice before pathology makes the final diagnosis.
93Thirdly, although beyond this discussion, it should be realised that the accuracy
94of the diagnosis by pathology cannot be established since it is the gold standard
95and thus cannot be compared with a better test. More fundamental is that
96Fourthly, accuracies and predictive values of abnormal or suspicious findings are
97crude estimates and should need to be stratified for factors such as for breast
98densities, which increases the difficulty of interpretation and the risk of
99abnormal findings (Freer, 2015, Schünemann et al., 2020). Finally, although
100beyond this discussion, it should be realised that the accuracy of the diagnosis
101by pathology cannot be established since it is the gold standard, which cannot be
102compared with a better test. In addition, it is unclear whether histopathological
103parameters reflect the expression profile or molecular genetic features (Simpson
104et al., 2005). ~~(Kubota et al., 2023)~~

105These poor predictive values of mammography and the stepwise clinical
106approach result in more lumpectomies and mastectomies ~~(RR 1.31, 95% CI 1.22-~~
107~~to 1.42)~~ (Gøtzsche and Jørgensen, 2013, Schünemann et al., 2020) with an
108estimated 30% overdiagnosis out of prudence. Without discussing whether
109screening results in a reduction in breast cancer mortality, we should realise-
110realise that a reduction of breast cancer mortality by 15% and an overdiagnosis

111and overtreatment of 30% results in preventing one mortality for every 2000
112women screened for ten years, but at the cost of 10 women being treated
113unnecessarily, and 200 women having experienced psychological distress and
114anxiety. Although the exact figures can be discussed, the conclusion of
115overdiagnosis and little effect on mortality seems repetitively confirmed
116(Bretthauer et al., 2023)

117**Breast cancer screening, clinical medicine and Bayesian statistics**

118Clinical medicine is multivariate, and in the individual woman (Koninckx et al.,
1192023), the probability of having breast cancer varies not only with the image of
120mammography (or MRI) but also with the clinical exam, ~~AND~~ age, ~~AND~~ heredity,
121~~AND~~ obesity, ~~AND~~ breast density, ~~AND~~ race and many other factors. The
122estimation of the probability in the individual woman that prediction of the risk of
123an abnormal or suspicious lesion is a cancer in the individual woman should
124include interpret the mammography findings considering all these other risk-
125factors besides the mammography. This requires multivariate Bayesian
126statistics, but these results are unfortunately not ~~yet~~ available. For example, since
127multivariate RCTs are difficult to perform, For example, the added value of
128mammography following a negative clinical exam, with having a reported
129sensitivity of 54% and specificity of 94% (Jatoi, 2003), is still unknown, ~~could~~
130~~not be demonstrated.~~ (Arun Kumar and Sasikala, 2023) Therefore, the
131translation of an abnormal or suspicious lesion into the management of the
132individual woman remains a clinical judgment. For example, the added value of
133mammography following a negative clinical exam, having a sensitivity of 54%
134and specificity of 94% (Jatoi, 2003), could not be demonstrated.
135This explains that recommendations for the diagnostic workup of abnormal or
136suspicious breast lesions vary vary from no recommendation to triple diagnosis
137by imaging, physical examination, and biopsy. (Jhangiani et al., 2023)

138 **Breast cancer screening and hormone replacement therapy**

139 Hormone replacement therapy (HRT) increases breast density and thus the risk
140 of finding abnormal and suspicious lesions (Rutter et al., 2001, Freer, 2015, Azam
141 et al., 2018). Therefore, it was suggested that HRT be stopped for a few months
142 before screening mammography is performed (Beckmann et al., 2013).

143 Unfortunately stopping HRT for several months is clinically poorly accepted.

144 Since breast cancer screening by mammography is recommended to be
145 performed only every two years, the risk of delaying the workup and eventual
146 treatment of abnormal or suspicious suspected lesions for 3 to 6 months must be
147 very low. Therefore, managing women on HRT and abnormal or suspicious lesions
148 on screening mammography must be clinically individualised.

149 Considering all risk factors, the increase of abnormal or suspicious images in
150 women taking HRT and the low risk of delaying therapy for a few months, it
151 may even be considered to stop HRT for 3 to 6 months before repeating

152 mammography. The women whose repeat mammography returns to normal will
153 have avoided unnecessary cancer workups and surgery, and the continuation of
154 HRT can be discussed. With GnRH antagonists becoming widely available, this
155 therapy might even be considered for a few months in women with dense
156 breasts before menopause.

157 **Conclusion and discussion**

158 The limitations of performing multivariate RCTs. The low prevalence of breast-
159 cancer and the frequentist statistical data-analysis hamper the interpretation of
160 breast cancer screening by mammography. It is still not fully appreciated that the
161 predictive value of any test is strongly affected by the prevalence of the disease
162 (Lesaffre and Lawson, 2012). For breast cancer with a prevalence of 0.5%, the
163 positive predictive value is only 16% for sensitivities of 67% and specificities of
164 98%. Unfortunately, negative predictive values cannot be estimated since the
165 women in whom breast cancer is missed (false negatives) are unknown since

166they have not been further explored. Also, although estimated as high as 40%
167(Ryser et al., 2022), the true incidence of false positives is not clear since
168expectant management of suspicious lesions is considered unethical. Clinically
169important is that frequentist statistical analysis calculates the positive predictive
170values of breast cancer screening but not the added value besides age, heredity
171and breast density. This would require a Bayesian statistical approach, which has
172not been performed yet to the best of our knowledge.

173The diagnostic workup of abnormal and suspicious lesions follows a clinical logic
174o, with repeat exams and, eventually, biopsies, with the dogma (but without
175data) that delaying the workup and treatment can only be harmful.
176Overtreatment and unnecessary fear have been discussed, but the hypothesis
177that some lesions could disappear spontaneously, especially in women stopping
178HRT, has not been tested.

179It is beyond this manuscript to discuss the limitations of frequentist statistics or
180the need to individualise therapy and to consider all predictive
181factors or medical and corporate pressures, as we recently did for endometriosis
182(Koninckx et al., 2022). However, individualising follow-up and
183postponing exams for three months in women with abnormal or suspicious
184lesions and dense breasts while HRT is stopped has the potential to decrease
185overtreatment. Although it would be an easy trial to demonstrate that some
186abnormal images become normal after stopping HRT for three months, it is close
187to impossible to demonstrate that a delay of 3 months does not increase the risk
188since proving the absence of an effect requires huge numbers. Therefore, an
189adequately performed RCT to demonstrate a decreased risk of unnecessary
190treatments without increasing risks is unlikely to be performed. In conclusion,
191individualising follow-up and letting the woman decide whether to stop HRT for
192three or more months before repeating mammography has the potential to

193 decrease overtreatment. This is an important message for clinicians involved in
194 the long-term follow-up of women after surgery for other indications such as
195 endometriosis.

196

197 Word count: 1196

198

199 **Disclosure:** None of the authors has a conflict of interest to declare.

200 **IRB board:** The IRB board of KU Leuven Gasthuisberg confirmed in writing that
201 for experiments not involving human or animal tissues, IRB board approval was
202 not requested

203 **Authors role.** All authors discussed the manuscript and agreed with the final
204 manuscript.

205 **Funding:** No funding

206

207 **List of legends**

208 Fig 1 Positive predictive values ($PPV = \frac{\text{sensitivity} \times \text{prevalence}}{\text{sensitivity} \times \text{prevalence} + (1 - \text{specificity})(1 - \text{prevalence})}$) of a mammography if
209 sensitivities and specificities are 73% and 88% or 80% and 98%. This
210 illustrates the crucial importance of the specificity and of the prevalence of the
211 disease (0.5 and 1% indicated).

213

214 Reference List

215

216 Arun Kumar S and Sasikala S. Review on Deep Learning-Based CAD Systems for
217 Breast Cancer Diagnosis. Technol. Cancer Res. Treat. 2023;22:15330338231177977.

219

220 Azam S, Lange T, Huynh S, et al. Hormone replacement therapy, mammographic
221 density, and breast cancer risk: a cohort study. Cancer Causes & Control.
222 2018;29:495-505.

223

224Beckmann KR, Farshid G, Roder DM, et al. Impact of hormone replacement
225therapy use on mammographic screening outcomes. *Cancer Causes & Control*.
2262013;24:1417-26.

227

228Belgium (2021). "Jaarrapport bevolkingsonderzoeken naar kanker 2021."
229Retrieved 03-092023, from [https://www.zorg-en-
230gezondheid.be/nieuws/jaarrapport-bevolkingsonderzoeken-naar-kanker-2021](https://www.zorg-en-gezondheid.be/nieuws/jaarrapport-bevolkingsonderzoeken-naar-kanker-2021).

231

232Bretthauer M, Wieszczy P, Løberg M, et al. Estimated Lifetime Gained With
233Cancer Screening Tests: A Meta-Analysis of Randomized Clinical Trials. *JAMA*
234Intern Med. 2023;

235

236Freer PE. Mammographic breast density: impact on breast cancer risk and
237implications for screening. *Radiographics*. 2015;35:302-15.

238

239Gao Y, Lin J, Zhou Y and Lin R. The application of traditional machine learning and
240deep learning techniques in mammography: a review. *Front Oncol*.
2412023;13:1213045.

242

243Gøtzsche PC and Jørgensen KJ. Screening for breast cancer with mammography.
244Cochrane Database of Systematic Reviews. 2013;

245

246Jatoi I. Screening clinical breast examination. *The Surgical clinics of North*
247America. 2003;83:789-801.

248

249Jhangiani N, Philip M and Jatoi I. Breast cancer screening guidelines:
250discrepancies raise concerns about validity. *Breast Cancer*. 2023;30:869-71.

251

252Koninckx PR, Ussia A, Alsuwaidi S, et al. Reconsidering evidence-based
253management of endometriosis. *Facts Views Vis Obgyn*. 2022;14:225-33.

254

255Koninckx PR, Wattiez A and Saridogan E. Statistics, truth finding and predictions:
256what every gynaecologist and researcher should know. *Facts Views Vis Obgyn*.
2572023;15:95-7.

258

259Kubota K, Nakashima K, Nakashima K, et al. The Japanese breast cancer society
260clinical practice guidelines for breast cancer screening and diagnosis, 2022
261edition. *Breast Cancer*. 2023;

262

263Lesaffre E and Lawson AB. *Bayesian biostatistics*. United Kingdom, John Wiley &
264Sons, ltd.2012.

265

266Mushlin AI, Kouides RW and Shapiro DE. Estimating the accuracy of screening
267mammography: a meta-analysis. *Am. J. Prev. Med*. 1998;14:143-53.

268

269Rutter CM, Mandelson MT, Laya MB, et al. Changes in breast density associated
270with initiation, discontinuation, and continuing use of hormone replacement
271therapy. JAMA. 2001;285:171-6.

272

273Ryser MD, Lange J, Inoue LY, et al. Estimation of Breast Cancer Overdiagnosis in
274a U.S. Breast Screening Cohort. Ann Intern Med. 2022;175:471-8.

275

276Schünemann HJ, Lerda D, Quinn C, et al. Breast Cancer Screening and Diagnosis:
277A Synopsis of the European Breast Guidelines. Ann Intern Med. 2020;172:46-56.

278

279Simpson PT, Reis-Filho JS, Gale T and Lakhani SR. Molecular evolution of breast
280cancer. The Journal of Pathology. 2005;205:248-54.

281

282