Decision theory for precision therapy of breast cancer

Michael Kenn1, Dan Cacsire Castillo-Tong2, Christian F. Singer2, Rudolf Karch1, Michael Cibena1, Heinz Koelbl3 & Wolfgang Schreiner1,*

Correctly estimating the hormone receptor status for estrogen (ER) and progesterone (PGR) is crucial for precision therapy of breast cancer. It is known that conventional diagnostics (immunohistochemistry, IHC) yields a significant rate of wrongly diagnosed receptor status. Here we demonstrate how Dempster Shafer decision Theory (DST) enhances diagnostic precision by adding information from gene expression. We downloaded data of 3753 breast cancer patients from Gene Expression Omnibus. Information from IHC and gene expression was fused according to DST, and the clinical criterion for receptor positivity was re-modelled along DST. Receptor status predicted according to DST was compared with conventional assessment via IHC and gene-expression, and deviations were flagged as questionable. The survival of questionable cases turned out significantly worse (Kaplan Meier \( p < 1\% \)) than for patients with receptor status confirmed by DST, indicating a substantial enhancement of diagnostic precision via DST. This study is not only relevant for precision medicine but also paves the way for introducing decision theory into OMICS data science.

Background and significance. Precision medicine relies on biomarkers for selecting the most adequate therapy for each single patient1–3. In particular, individually optimized treatment of breast cancer patients has widely been studied, drawing on molecular subtyping4–7, networks and pathways8–12 or more general gene expression signatures13–20. Among such signatures, PAM5021,22, PREDICT23, the Gene expression Grade Index24 and the Genomic Grade Index25 have been widely recognized. For a survey see Huang, C. C. et al.26.

Hormone receptor status (estrogen, ER and progesterone, PGR) as well as the human epidermal growth factor 2 receptor (HER2) are considered most important27,28 and are estimated by immunohistochemistry (IHC) in clinical practice29–36. If at least one of both receptors (ER or PGR) is found positive (i.e. receptors are present), hormone therapy will suffice for an effective treatment, avoiding the much more severe side effects of chemotherapy. However, if the receptor status should accidentally be estimated falsely positive, hormone treatment will not work and the patient may be deprived from life-saving chemotherapy.

Due to its focal importance, receptor assessment quality has been scrutinized in multicenter studies37,38, suggesting up to 20% misclassifications39–41. Guidelines have been issued39,42,43 but improving assessment precision remains a key issue44. A most natural way to increase reliability is to enhance IHC-estimates by additional information, e.g. from gene expression. Methodological advances were reported for data normalization45–48 as well as for gene expression analysis49–51. A survey was given by Bolstad, Irizarry, Astrand and Speed52. Chen modelled the probability distributions for gene-expressions of ER, PGR and HER2 and set cutoff values at a posterior probability of 0.5 to discriminate receptor positive from negative patients53. Likewise, Laas set cutoffs for receptor positivity within the frequency distributions of gene-expression of ER and HER241. Gong used random sampling to estimate optimum thresholds for ER and HER2 and verified these in a test cohort54. Bergqvist set thresholds for HER2 expression according to visual inspection55. Lopez introduced fuzzy rules to identify breast cancer biomarkers56. Concluding, prediction algorithms for low dimensional input space (such as hormone receptors), yield fairly similar results. In previous papers57,58 we have worked on such approaches, applying standard statistical means (odds-products). In this work we expand the approach by drawing on Dempster Shafer decision Theory (DST)59.

DST has been widely applied in self driving cars60–62, driver’s vigilance monitoring61, aircraft technology63,64 and also in medical settings, e.g. image based decisions65, diagnosis of prostate66 and breast cancer67. In all these...
applications, unclear or even contradicting information from several sources are fused to arrive at decisions of optimized precision.

Using decision theory for receptor status assessment. Classical probability theory characterizes the chance for an event to occur by a single number, its probability. A probability, \( p \), characterizes the event as such, including its circumstances, but disregards the measuring process itself. The performance of the measuring process is characterized by sensitivity and specificity, in short by its receiver operating characteristics (ROC). In contrast, DST includes the measuring process and characterizes expected outcomes (say ‘receptor positive’, ‘Rez’) by two numbers, the belief \( \text{bel}(\text{Rez}^+) \) and the plausibility, \( \text{pl}(\text{Rez}^+) \), together called ‘evidence’ of the outcome. Note that this definition of evidence specifically relates to the framework of decision theory and differs from more general meanings of the term, e.g. in ‘evidence based medicine’. The belief indicates the chance to correctly measure ‘positive’ by virtue and quality of the measuring process. The plausibility indicates the chance that the reading ‘positive’ (for an actually positive receptor) could also result (a) by chance or (b) represent a false positive outcome (for a receptor status truly ‘negative’). For each measurement with dichotomic outcome, the belief \( \text{bel}(\text{Rez}^+) \) and the plausibility \( \text{pl}(\text{Rez}^+) \) are explained in the next section (Eqs. 4–6).

Materials and methods

Data curated and used. In order to establish a comprehensive database\(^6\), we decided to re-use published gene expression data\(^8\) and screened the Gene Expression Omnibus (GEO)\(^5\) for breast cancer studies using the Affymetrix chip U133A + 2.0. We found and curated 38 studies with 3753 samples.

Out of numerous methods available for normalization\(^46,73–76\) as compared by Bolstad\(^52\) and evaluated in our previous work\(^77\), we used RMA (MATLAB affyirma)\(^8\) and standardized data for each sample. Further batch corrections have been scrutinized in our previous work\(^77\) and were found ambiguous. We therefore refrained from performing them.

We adopted receptor genes and selected co-genes (see Table 1) from our previous paper\(^57\). Whereas in multiple logistic regression all genes would be incorporated simultaneously in the predictor (multiple logistic regression), the ODDS method performs logistic regressions on single genes and then combines the odds. In other words, a gene predictor would be given by the sum of \( \text{odds}(\text{Rez}^+) \), that is, the sum of \( \text{odds}(\text{Rez}^+) \) for all genes. DST additionally accommodates uncertainty, \( \theta \), defined as \( \theta(\text{Rez}^+) = \text{pl}(\text{Rez}^+) - \text{bel}(\text{Rez}^+) = 1 - \text{bel}(\text{Rez}^-) - \text{bel}(\text{Rez}^+) \), i.e. the uncertainty of an event is the difference between its plausibility and its belief. For brief notation we use \( \text{bel}(\text{Rez}^+) = \alpha \) and \( \text{bel}(\text{Rez}^-) = \beta \).

The main parts of this paper address readers familiar with DST and demonstrate how to combine several evidences for single hormone receptors and further combine several receptor estimates towards an overall hormone receptor status. Readers not yet familiar with DST may first see the ‘supplementary methods’ and then resume to read the following chapters.

| Probe set | Logistic regression parameters | Logistic regression quality | Upper limits for beliefs |
|-----------|-------------------------------|----------------------------|-------------------------|
| Gene      | ESR1                          | 205225_at                  | 3.798 ~ 2.302           | 1106.4 2227 | 0.820 0.885 |
| Co-gene   | AGR3                          | 228241_at                  | 1.528 ~ 1.557           | 1254.1 0.790 0.841 |
| Gene      | PGR                           | 208305_at                  | 0.886 ~ 1.944           | 1329.3 1714 0.755 0.660 |
| Co-gene   | ESR1                          | 205225_at                  | 4.112 ~ 2.422           | 845.0 0.826 0.878 |

Table 1. Receptor genes, co-genes and parameters from logistic regression. Probe sets refer to the Affymetrix chip U133A + 2.0. For ‘deviance of fit’, see p.118 in McCullagh\(^81\). Note the double role of ESR1: It is the very receptor gene for ER but also serves as best co-gene for PGR, see our previous work\(^57\) regarding the selection of co-genes. Upper limits for beliefs (\( \alpha, \beta \)) are explained in the next section (Eqs. 4–6).
imputation we used default parameters and the threshold $|\text{score}| > \log ((1 - \varepsilon) / \varepsilon)$ with $\varepsilon = 0.15$, like in our previous work. Finally, $1798 + 1010 = 2808$ samples remained as HER2− according to the sODDS method, see Fig. 1.

These 2808 samples were then checked regarding ER and PGR as follows: If IHC-estimates were available for ER as well as PGR, samples were taken as is (1714 samples). If IHC was only available for ER and the status of PGR could be safely imputed via ODDS with sufficient score (see above), samples were also included (513 samples). If IHC was missing for both, ER as well as PGR, samples were taken only if both status could be imputed safely via ODDS (332 samples). Thus, $1714 + 513 + 332 = 2559$ samples were retained with receptor status established most accurately by state of the art (IHC plus ODDS $\equiv$ sODDS), see Fig. 2. We used only these to evaluate the benefit of applying decision theory.

Survival rates and quality of previous receptor status assessments. A study-wise evaluation of receptor status assessment via IHC and ODDS was performed, see Supp. Table 2. The concordance of receptor status between IHC and ODDS is much better for ER than for PGR, due to larger variance of ER in gene expression data. Results were further elaborated in a control chart, see Fig. 3. The dotted line represents $p$, the overall average rate of differences, computed from Supp. Table 2 as $p = 159 / 2227 \approx 0.071$. Markers within bars represent expectation values $(\Delta \text{IHC, ODDS} + 1)/(N_{\text{IHC}} + 2)$ of discordance rates derived from the beta-distribution, see the mathematical formalism in section ‘Concordance between IHC and gene expression’ in supplementary materials. Bars denote 95% confidence intervals of the respective mean values, allowing the qualification of studies against each other: If an upper bound lies below the dotted line in Fig. 3, the respective study has a discrepancy rate significantly below average (green). If a lower bound lies above the dotted line, the discrepancy rate of that study is significantly above average (red).

Similarly, a control chart for progesterone is shown in Supp. Fig. 2.

Additionally, we evaluated the scmgene-marker (from package genefu) for estrogen only, since scmgene does not give PGR estimates. Note that scmgene was evaluated for all (3753) samples but results were only compared for those 2559 with HER2− and reliable hormone receptor status, see Supp. Table 1.

Survival has been related to receptor status for ER (Supp. Fig. 3), PGR (Supp. Fig. 4) and hormone overall, i.e. either estrogen or progesterone being positive, see Fig. 4.

Responsibility functions for gene expression. In order to compute DST evidences from gene expression, we will first define responsibility functions as a prerequisite. The concept was coined by Hastie for weighing two distributions against each other regarding membership of a measurement in question. We use this for distributions of gene-expression data and their indication of receptor status. Based on these, ‘mass functions’ will be computed, and these converted into evidence:

Gene expression $\rightarrow$ responsibility functions $\rightarrow$ mass functions $\rightarrow$ evidence.

In the main part of this paper for brevity we skip the concept of masses (hence above shown in parenthesis) and present formulae for evidence directly based on responsibility functions. Masses are relegated to supplementary methods.
Responsibility functions are exemplified now for the expression of one gene, labelled ‘Expr’, to keep notation slim. It applies analogously to the other genes considered.

We construct the responsibility function, \( r_+ \), for receptor positivity from logistic regression (fitting coefficients: \( c_0, c_1 \)) of IHC receptor status versus gene expression, \( x_{\text{Expr}} \), see Eq. (1) and the increasing dotted red line in Fig. 5. The (dual) responsibility function, \( r_- \), for receptor-negativity is simply given by \( r_- = 1 - r_+ \), see Eq. (1) and the decreasing blue dotted line in Fig. 5.

Logistic regression is carried out separately for estrogen receptor gene (ESR1), estrogen co-gene (AGR3), progesterone receptor gene (PGR) and its co-gene (ESR1). Results are consolidated in Table 1.

Obtaining evidences. We define DST masses (see supplementary methods) and evidences by drawing on responsibility functions as follows: We observe in Fig. 5 that \( r_+ \to 1 \) as gene expression approaches its maximum. However, not even maximum gene expression in reality provides total certainty of receptor status being positive. To account for this fact, the responsibility function has to be multiplied by an upper bound, \( \hat{\alpha}_{\text{Expr}} \), for the belief in ‘+’. Similarly \( \hat{\beta}_{\text{Expr}} \) acts as upper bound for the belief in ‘−’. Next we show how \( \hat{\alpha}_{\text{Expr}} \) and \( \hat{\beta}_{\text{Expr}} \) are obtained, see Fig. 6.

Out of all positive IHC-measurements (represented by the interval (0,1)), a certain fraction has been confirmed positive by gene expression, due to virtue of the method. This fraction of measurements is represented by the interval \((0,\hat{\alpha}_{\text{Expr}})\) as part of the interval (0,1). Given a positive IHC measurement, the plausibility \( \hat{\beta}_{\text{Expr}} = 0 \), and hence the rest of the interval \((0,1)\) represents nothing but uncertainty according to DST: \( \theta_{\text{Expr}} = 1 - \hat{\alpha}_{\text{Expr}} - \hat{\beta}_{\text{Expr}} = 0 \). As depicted in Fig. 6, some fraction, \( \theta_{\text{Expr}} \), of uncertain predictions from gene expression will result positive merely by chance, not by virtue of the method. Both parts together represent all true positive

---

**Figure 2.** Selection of samples with definite hormone receptor status. Selection started from 2802 samples HER2−, see Fig. 1. If an IHC estimate for estrogen (ER) was available, ER was considered defined (ER def., 2365). Otherwise the ODDS method37 was applied, yielding either a safe result (imp. ER, 414) or resulting uncertain (imp. ER unc.), contributing to the set of 249 samples shown in beige. ODDS was applied to the set of 414 samples to impute progesterone status, either safely (imp. PGR, 332, green) or ending up uncertain (imp. PGR unc.), contributing to the beige set of 249. Finally, out of the 2365 samples with IHC available, IHC for progesterone was either available (IHC PGR def., 1714, green), safely imputed by ODDS (imp. PGR, 513, green) or uncertain (imp. PGR unc.), again contributing to the beige set. All in all, IHC receptor status was available for 1714 samples and safely imputed for 513 + 332 = 845 samples. This resulted in 2559 samples with definite hormone receptor status, out of which 1432 were positive (red) and 1127 negative (blue). These constituted the input set for DST, as shown in Fig. 7. Note that no samples exist with PGR\textsubscript{IHC} defined and ER\textsubscript{IHC} imputed.
IHC-measurements: $\alpha_{\text{Expr}} + \theta_{\text{Expr}}^{+} = TP$. A second part of uncertainty, $\theta_{\text{Expr}}^{-}$, will represent erroneously assessed false positives ($FP$), although being truly negative samples. Based on this nomenclature we proceed as follows:

Consider all samples qualified ‘uncertain’ by gene expression, see Fig. 6. The part $\theta_{\text{Expr}}^{+}$ represents truly positive samples, whereas $\theta_{\text{Expr}}^{-}$ are truly negative ones. We may now assume on good grounds that the ratio

![Figure 3](https://www.nature.com/scientificreports/)  
**Figure 3.** Control chart for estrogen receptor status: differences between IHC and ODDS gene expression estimates. x-axis: label of study. y-axis: rate of difference. Dotted line: overall mean rate of differences between IHC and ODDS $p = 159/2227 = 0.071$, see SuppTable 2. Markers (x) within error bars: rate of discordant samples in respective study. Lower and upper bounds of error bars denote 95% confidence limits. If the upper bound lies below the dotted line, the respective study has a discrepancy rate significantly below average (green). If the lower bound lies above the dotted line, the respective study has a discrepancy rate significantly above average (red).

![Figure 4](https://www.nature.com/scientificreports/)  
**Figure 4.** Survival free from recurrence for hormone status. Hormon status is considered positive, if either ER or PGR are positive, otherwise negative. Kaplan Meier estimates of survival free from recurrence for 870 patients (who had rfs-data) for different hormone receptor status. Log-rank test $p = 4.2e^{-06}$.
θ + Expr / θ − Expr among the uncertain ones from gene expression is similar (equal) to the ratio of positives and negatives found by IHC, see also Shoyaib82. Considering the contingency table of IHC-measurements, note that positive measurements comprise \( TP + FN \), whereas negative ones are composed as \( TN + FP \). Hence we may put

\[
\frac{\theta^+_{\text{Expr}}}{\theta^-_{\text{Expr}}} = \frac{TP + FN}{TN + FP}
\]

Moreover, the fraction of gene expression measurements being positive by virtue or chance, \( \alpha_{\text{Expr}} + \theta^+_{\text{Expr}} \), is set equal to the fraction of IHC-true positives:

\[
\alpha_{\text{Expr}} + \theta^+_{\text{Expr}} = \frac{TP}{TP + FP}
\]

Since \( \alpha_{\text{Expr}} + \theta^+_{\text{Expr}} + \theta^-_{\text{Expr}} = 1 \), Eqs. (2 and 3) can now be solved to yield the maximum possible belief in positive receptor status (derived from the gene in question):

\[
\alpha_{\text{Expr}} = \frac{TP \cdot TN - FP \cdot FN}{(TP + FP) \cdot (TN + FP)}
\]

Likewise we obtain for the maximum belief in negative status:

\[
\beta_{\text{Expr}} = \frac{TP \cdot TN - FN \cdot FP}{(TN + FN) \cdot (TP + FN)}
\]

Considering these maximum beliefs, we finally obtain for the evidence derived from gene expression, for illustration see Fig. 5:

**Figure 5.** Decision theory evidences obtained from logistic regression. A logistic regression of IHC receptor status (IHC+ ≡ 1, IHC− ≡ 0) versus gene expression (\( x_{\text{Expr}} \)) was performed to obtain the responsibility function for receptor positivity \( r_+ \) (dotted red curve) and \( r_- \) (dotted blue). It will be shown later (Eq. 6) that \( r_+ \) has to be multiplied by an upper limit, \( \alpha_{\text{IHC}} \), to obtain the actual belief \( \alpha_{\text{Expr}}(x_{\text{Expr}}) = \alpha_{\text{IHC}} \cdot r_+(x_{\text{Expr}}) \), see the solid red curve. Likewise \( \beta_{\text{Expr}}(x_{\text{Expr}}) = \beta_{\text{IHC}} \cdot r_-(x_{\text{Expr}}) \) (solid blue). Uncertainty: ochre. For a given expression value, e.g. \( x_{\text{Expr}} = 2 \), one can read off belief in positive (\( \alpha \)), belief in negative (\( \beta \)) and uncertainty (\( \theta \)).

**Figure 6.** Estimating an upper bound for belief in receptor positivity. True positives (TP) and false positives (FP) can be used to derive a first equation in \( \alpha_{\text{Expr}} \) and \( \theta_{\text{Expr}} \) (Eq. 3). A second equation may be deduced by including TN and FN, Eq. (2). Both can be solved to estimate \( \alpha_{\text{Expr}} \) and \( \theta_{\text{Expr}} \). From these follow upper bounds for the belief in positivity (\( \hat{\alpha}_{\text{Expr}} \)) and negativity (\( \hat{\beta}_{\text{Expr}} \)), Eqs. (4 and 5), respectively.
\[ a_{Exp} (x_{Exp}) = a_{Exp} \cdot r_+ (x_{Exp}|c_0, c_1) \]
\[ b_{Exp} (x_{Exp}) = b_{Exp} \cdot r_- (x_{Exp}|c_0, c_1) \]

(6)

Note that the whole formalism (Eqs. 1–6) is performed separately for each gene (and co-gene), using the data of corresponding IHC measurements. For generality and brevity of above notation, the subscript ‘Expr’ stands for each of those genes.

**Completing the decision theory framework.** With responsibility functions and evidences available we may now assemble the decision framework as follows.

Each hormone receptor (ER, PGR) is assessed via three sources of information, each yielding separate evidence (in our case of dichotomous data: 2 numbers):

1. IHC → evidence \( (\alpha_{IHC}, \beta_{IHC}) \)
2. gene expression of the receptor gene \( \rightarrow (\alpha_{Gen}, \beta_{Gen}) \)
3. gene expression of a co-gene \( \rightarrow (\alpha_{Co}, \beta_{Co}) \)

In a first step, evidences for the receptor gene \( (\alpha_{Gen}, \beta_{Gen}) \) and its co-gene \( (\alpha_{Co}, \beta_{Co}) \) are combined by the evidence combination rule (ECR) introduced by Dempster\(^8\) (labelled ‘⊕\(_D\)’) to obtain joint evidence \( (\alpha_{Exp}, \beta_{Exp}) \) from gene expression:

\[ (\alpha_{Exp}, \beta_{Exp}) = (\alpha_{Gen}, \beta_{Gen}) \oplus_D (\alpha_{Co}, \beta_{Co}) \]

(7)

As explained in supplementary methods, the operation \( \oplus_D \) yields in detail:

\[ \alpha_{Exp} = \frac{\alpha_{Gen}\alpha_{Co} + \theta_{Gen}\alpha_{Co} + \alpha_{Gen}\theta_{Co}}{1 - \alpha_{Gen}\beta_{Co} - \beta_{Gen}\alpha_{Co}} \]
\[ \beta_{Exp} = \frac{\beta_{Gen}\beta_{Co} + \theta_{Gen}\beta_{Co} + \beta_{Gen}\theta_{Co}}{1 - \alpha_{Gen}\beta_{Co} - \beta_{Gen}\alpha_{Co}} \]
\[ \theta_{Exp} = 1 - \alpha_{Exp} - \beta_{Exp} = \frac{\theta_{Gen}\theta_{Co}}{1 - \alpha_{Gen}\beta_{Co} - \beta_{Gen}\alpha_{Co}} \]

(8)

This evidence for gene expression is—in a second step—combined with the evidence from IHC, \( (\alpha_{IHC}, \beta_{IHC}) \), either using rule \( \oplus_Y \) once more or, alternatively the ECR defined by Yager\(^9\), labelled \( \oplus_Y \). Since \( \oplus_Y \) more easily accommodates contradicting evidence (from IHC and gene expression), we opt for \( \oplus_Y \) and obtain:

\[ \alpha_{Rez} = \alpha_{Exp}a_{IHC} + \theta_{Exp}a_{IHC} + \alpha_{Exp}\theta_{IHC} \]
\[ \beta_{Rez} = \beta_{Exp}b_{IHC} + \theta_{Exp}b_{IHC} + \beta_{Exp}\theta_{IHC} \]
\[ \theta_{Rez} = \theta_{Exp}b_{IHC} + \alpha_{Exp}b_{IHC} + \beta_{Exp}\alpha_{IHC} \]

(9)

In supplementary methods we explain in detail how general concepts of DST (capable of handling a set of multiple outcomes) boil down to Eqs. (8 and 9) in case of just two outcomes, receptor status \( \{\text{pos, neg}\} \).

The above procedure is carried out similarly to obtain the combined evidence for the PGR-status \( (\alpha_{PGR}, \beta_{PGR}) \). Note that notation is turned from general to specific in the following \( (\alpha_{Rez} \rightarrow \alpha_{ER} \text{ and } \alpha_{Rez} \rightarrow \alpha_{PGR}, \text{ respectively}) \).

In a last step, the clinical decision for ‘hormone versus chemo’ is modelled along the lines of DST. Clinically, a patient is considered receptor positive (and will receive hormone therapy) if either ER or PGR (or both) are positive. Classically, this is a crisp, logical decision as stated. DST however, allows more elaborate combination rules to combine evidences for estrogen \( (\alpha_{ER}, \beta_{ER}) \) and progesterone \( (\alpha_{PGR}, \beta_{PGR}) \).

Thus, considering ER as well as PGR, each being assessed by IHC, gene and co-gene, performing some algebra (detailed in supplementary methods) yields the overall evidence for ‘hormone receptor status’ \( \alpha_H = \beta_H (\text{pos}), \beta_H = \beta_H (\text{neg}) \):

\[ \alpha_H = \alpha_{ESR} + \alpha_{PGR} - \alpha_{ESR}\alpha_{PGR} \]
\[ \beta_H = \beta_{ESR} \cdot \beta_{PGR} \]

(10)

As always we have to accept an amount of uncertainty given by \( \theta_H = 1 - \alpha_H - \beta_H \), indicating hormone receptor status being indeterminable. From these evidences, a most reasonable decision rule can be derived:

| Hormone receptor positive | \( \alpha_H > \beta_H + \theta_H \) |
| Hormone receptor negative | \( \beta_H > \alpha_H + \theta_H \) |
| Hormone receptor indeterminable | \( |\alpha_H - \beta_H| \leq \theta_H \) |

(11)

In our simple case with only two outcomes these criteria reduce to \( \alpha_H > 0.5 \) for definitely receptor positive and \( \beta_H > 0.5 \) for definitely receptor negative.
Results

DST results for patient cohort. In our previous work, we considered receptors (ER, PGR) separately. For each of them routine estimates (based on IHC only, 1714 samples, see Fig. 2) could be significantly improved by adding information from gene expression via conventional statistics (odds ratio) and we label this former method 'ODDS'. Note that for patients lacking one or both IHC estimates, receptor status was imputed from gene expression according to ODDS, including an uncertainty region, as in our previous paper. However, to obtain a dataset of optimum quality for the present work, only samples rendering safe imputations were retained (513 + 332 = 845), see Fig. 2.

Then we demonstrated that further improvement in receptor assessment can be achieved by applying decision theory. Besides yielding new estimates for each receptor, DST allows to combine estimates for ER and PGR into a joined 'hormone' receptor estimate. Based on Eqs. (10 and 11), hormone receptor status was newly predicted for each patient, see the contingency Table 2 and Fig. 7.

To demonstrate the clinical benefit of DST, we consider two groups:

1. Patients diagnosed receptor positive by sODDS, and being confirmed by DST. They have correctly received hormone treatment, see the flow labelled 'a' in Fig. 7.
2. Patients diagnosed receptor positive by sODDS, but being questioned by DST, see flow 'b'. They have very probably also received hormone treatment but it remained ineffective. At the same time they might have been deprived of life-saving chemo.

Figure 8 shows a strikingly worse survival of patients assumed receptor positive although being negative (logrank–Wilcoxon \( p = 0.009 \)). Note that out of 1432 patients considered positive by sODDS, only for 651 survival data were available. Among these, 26 were questioned by DST. Nonetheless the difference was statistically significant at the 1%-level.

| Method DST | − | Uncertain | + | Σ |
|------------|---|-----------|---|---|
| Method sODDS | 1023 | 103 | 1 | 1127 |
| | 0 | 50 | 1382 | 1432 |
| Σ | 1023 | 153 | 1383 | 2559 |

Table 2. Receptor status due to decision theory (DST) compared with previous method (ODDS). 2559 Patients have been assessed for hormone receptor status by method sODDS to comprise the input dataset. DST was then applied to this dataset. A sample was considered H+ if at least one receptor was positive. Difference between methods sODDS and DST is reflected in Cohen’s \( \kappa = 0.88 \). Note that uncertain results were considered as separate category in computing \( \kappa \).

Figure 7. Alluvial flow diagram for patients originally diagnosed by ODDS and then checked by DST. Left margin: Hormone receptor assessed by method sODDS: positive (H+) and negative (H−). A sample was considered H+ if at least one receptor was positive. Lacking IHC estimates were imputed by ODDS and included only if imputation was safe, otherwise discarded. Hence, no patients with receptor status ‘undetermined’ appear in dataset subjected to DST. Right margin: Evaluation by DST confirms the vast majority of estimates from sODDS but leaves some 153 as undetermined (ochre colored). Only one patient is definitely considered false negative and changed to positive.
It is interesting to see what DST achieves as compared to ODDS when starting from the same patient cohort. We therefore took the same sample of patients (2559) as in Table 2 and Fig. 7, with 1432 receptor status positive and 1127 negative according to sODDS. Instead of DST we this time applied the ODDS method, including a uncertainty region, see Table 3.

Most strikingly, DST had flagged almost double as many patients ‘uncertain’ (156, see Table 2) as compared to ODDS (82), although the very same safety threshold was used. This clearly reflects an advantage of DST, since patients with questionable status would be re-evaluated before assigning the adequate therapy.

To characterize these differences in judgement we performed a face-to-face comparison between DST and ODDS on the very same input dataset, see Table 4. Samples additionally labelled ‘uncertain’ by DST originate to almost equal parts from negative (44) and positive (39) estimates according to ODDS. Only very few samples considered uncertain by ODDS, migrate to negative (1) or positive (11) according to DST.

Table 3. Receptor status according to ODDS. 2559 Patients have been assessed for hormone receptor status by sODDS to comprise the input dataset. A sample was considered $H^+$ if at least one receptor was positive. Difference in methods is reflected in Cohen’s $\kappa = 0.88$.

| Method sODDS | $-$ | Uncertain | $+$ | $\Sigma$ |
|--------------|----|-----------|----|--------|
| $-$          | 1035 | 52 | 40 | 1127 |
| $+$          | 31 | 30 | 1371 | 1432 |
| $\Sigma$     | 1066 | 82 | 1411 | 2559 |

Table 4. DST Decision theory compared to ODDS. The same dataset (2559 samples) was subjected to DST and ODDS, both including the same uncertainty regions as described. DST is seen to flag 153 patients as uncertain, i.e. of questionable hormone receptor status. As compared to ODDS almost double as many patients are detected to be re-checked before therapies are assigned. Difference in methods is reflected in Cohen’s $\kappa = 0.88$.

| Method DST | $-$ | uncertain | $+$ | $\Sigma$ |
|------------|----|-----------|----|--------|
| $-$        | 1022 | 44 | 0 | 1066 |
| Uncertain  | 1 | 70 | 11 | 82 |
| $+$        | 0 | 39 | 1372 | 1411 |
| $\Sigma$   | 1023 | 153 | 1383 | 2559 |

DST compared to ODDS. It is interesting to see what DST achieves as compared to ODDS when starting from the same patient cohort. We therefore took the same sample of patients (2559) as in Table 2 and Fig. 7, with 1432 receptor status positive and 1127 negative according to sODDS. Instead of DST we this time applied the ODDS method, including a uncertainty region, see Table 3.

Most strikingly, DST had flagged almost double as many patients ‘uncertain’ (156, see Table 2) as compared to ODDS (82), although the very same safety threshold was used. This clearly reflects an advantage of DST, since patients with questionable status would be re-evaluated before assigning the adequate therapy.

To characterize these differences in judgement we performed a face-to-face comparison between DST and ODDS on the very same input dataset, see Table 4. Samples additionally labelled ‘uncertain’ by DST originate to almost equal parts from negative (44) and positive (39) estimates according to ODDS. Only very few samples considered uncertain by ODDS, migrate to negative (1) or positive (11) according to DST.

Figure 8. Precise versus inadequate treatment based on hormone receptor status. Survival free from recurrence, according to Kaplan–Meier. Red: Patients treated correctly as receptor positive (flow (a) in Fig. 7). Ochre: Patients erroneously assigned positive (flow (b) in Fig. 7).
and Note that all points lie in one plane due to the definition (α chance equals the fraction of IHC-true positives (Eq. 3). Likewise we assumed that the ratio of uncertain positive convincing view on the classification, see Fig. 9. From the definition of DST decision, instead of just a single number as conventional probability theory does. In addition, information from different sources is joined according to mathematical rules rather than ‘clinical intuition’.

DST originates from elaborate evidence joining: Simple and/or combinations (such as estrogen-or-progesterone logics: ‘If ER or PGR receptor estimates is/are positive, the patient is considered positive’. In part, the strength of DST flags about twice as many patients as ‘uncertain’ (153) as ODDS (82). Additional patients subjected to a re-evaluation of hormone receptor status will improve precision of therapy.

For joining evidences from 2 genes for the same receptor we chose the Dempster ECR. For adding IHC and negative cases equals the ratio of positives and negatives (Eq. 2).

Figure 9. Decision space of hormone receptor status according to DST. (A) The three evidences of receptor status, belief in positive (αH), belief in negative (βH) and uncertainty (θH), of all samples shown in a three dimensional plot. Colors indicate receptor assignment according to DST (positive Δ red, negative Δ blue, uncertain Δ ochre). Dots: assignment via available IHC. Open circles: due to lacking IHC, assignment by imputation via DST. Decision borders according to Eq. (11) are indicated by dotted lines at αH = 0.5, βH = 0.5 and θH = 0.5. Samples with same status assigned appear confined within the same respective tetrahedral space. Note that all points lie in one plane due to the definition (αH + βH + θH = 1, see SuppEq. 16) within a triangle (surrounded by heavy lines). (B) The triangle plane in panel (A) seen from a slightly different aspect, along a vector parallel to plane (αH, βH), and perpendicular to the baseline of the triangle. Decision borders are more clearly visible in this aspect. In this projection the point (αH = 1, βH = 1) in the foreground of panel (A) coincides with the baseline of the triangle, and so do the axes αH and βH. Note that they run in opposite directions. Hormone receptor positive samples lie exclusively in the lower right triangle, negative ones in the lower left whereas uncertain results in the kite-shaped area.

Discussion

Dempster Shafer decision theory is a very potent framework. Its strength lies in combining information from several sources about the same item (here exemplified by IHC, Gene and Co-Gene estimates for one receptor) and also evidences for several different items, interacting in a biological setting (here exemplified by estrogen and progesterone regarding precision therapy of breast cancer).

We have started from some best methodology available up to now (sODDS), i.e. receptor status assessment via IHC enhanced by gene expression. ODDS had already been shown to improve precision therapy57, and here we demonstrate that further improvement is possible by application of DST. Since absolute truth of receptor status is unknown, the usefulness of our approach is demonstrated by disease free survival being severely degraded in patients flagged by DST as possibly wrongly diagnosed and treated. Explicit data on treatment are rare in the breast cancer studies downloaded from GEO. Hence we performed our calculations on the reasonable assumption that treatment was given according to the best knowledge available about receptor status.

The potency of DST is further underpinned by directly comparing both methods applied to the same dataset: DST flags about twice as many patients as ’uncertain’ (153) as ODDS (82). Additional patients subjected to a re-evaluation of hormone receptor status will improve precision of therapy.

Formulae and results derived therefrom incorporate three assumptions which we think were reasonable:

To compute basic belief assignments for each patient, we obtained DST mass functions by logistic regression. It proved more stable than other possible methods, e.g. fitting bimodal distributions such as Gaussian mixtures, as done in a previous study98.

For joining evidences from 2 genes for the same receptor we chose the Dempster ECR. For adding IHC evidence to those from gene expression, we opted for the Yager ECR.

To obtain upper limits α, β for belief we reasonably assumed that the fraction of positive results by virtue or chance equals the fraction of IHC-true positives (Eq. 3). Likewise we assumed that the ratio of uncertain positive and negative cases equals the ratio of positives and negatives (Eq. 2).

It is interesting to note that up to now clinical decisions follow the conventional (Cantor) non-exclusive OR-logics: ‘If ER or PGR receptor estimates is/are positive, the patient is considered positive’. In part, the strength of DST originates from elaborate evidence joining: Simple and/or combinations (such as estrogen-or-progesterone receptor positivity) are fundamentally refined, since DST draws on two numbers (belief and plausibility) in each decision, instead of just a single number as conventional probability theory does. In addition, information from different sources is joined according to mathematical rules rather than ‘clinical intuition’.

Joining evidences from both receptors in a more intricate way according to DST, allows for an intuitively convincing view on the classification, see Fig. 9. From the definition of θH = 1–αH–βH follows αH + βH + θH = 1 and, accordingly, DST beliefs (αH, βH, θH) of all samples lie in one single plane in 3-dimensions, cutting through the axes at αH = 1, βH = 1 and θH = 1, see panel A. Decision boundaries (dotted lines in Fig. 9) originate at 0.5 on each axis of evidence (αH = 0.5, βH = 0.5) and confine tetrahedrons, each containing exclusively positive (red) or negative (blue) samples, respectively. These tetrahedrons appear as equilateral triangles in the view shown.
in panel B. Uncertain samples (shown in ochre) appear within the pentahedron (panel A), which reduces to a kite-shaped area in the view shown in panel B.

While in this paper we have presented a very simple example to demonstrate feasibility and benefit of the method, possible applications are wide in scope. Pathology and laboratory medicine in many cases have to deliver crisp decisions even if measurements by complementary methods lack agreement. DST is the method of choice to handle such situations.

The final evidence obtained by DST also allows assigning weights (loss functions) to certain outcomes in order to model clinical consequences (Eq. 11). Suppose, for example, a false positive outcome entails much more detrimental consequences than a false negative does.

Several different methods are available for joining: The ‘Dempster additivity rule’, works best with rather concordant variables. Needless to mention that correlation must not be too close to unity, as in this case the second variable would not add any significant information to the first one. As opposed, the ‘Yager rule’, lends itself to join evidences from rather contradicting sources.

OMICS data often comprise a considerable number of estimates, say counts for multiple fragments out of a gene, needing to be condensed to characterize the gene (activity) as a whole. A plethora of similar situations can rigorously be handled by DST. Even if some of these estimates should be (partly) contradictive, DST is an appropriate tool for consolidating. Formulae need to be adapted to the particular situation in question.

We think that this work may help pave the way for decision theory entering OMICS data science.
30. Harris, L. N.
31. Singer, C. F.
32. Harbeck, N. & Gnant, M. Breast cancer.
33. Wirapati, P.
35. Regan, M. M.
37. Bartlett, J. M.
38. Lee, M., Lee, C. S. & Tan, P. H. Hormone receptor expression in breast cancer: postanalytical issues.
39. Hammond, M. E., Hayes, D. F., Wolff, A. C., Mangu, P. B. & Temin, S. American society of clinical oncology/college of American pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer.
44. Li, Q.
46. Wu, Z. A review of statistical methods for preprocessing oligonucleotide microarrays.
47. Wu, Z. & Irizarry, R. A. Preprocessing of oligonucleotide array data.
55. Bergqvist, J.
59. Gordon, J. & Shortliffe, E. H. in Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project (eds. Buchanan, B. G. & Shortliffe, E. H.) Chap. 13, 832–838 (Addison-Wesley Publishing, 1984).
60. Högger, A. Dempster Shafer Sensor Fusion for Autonomously Driving Vehicles : Association Free Tracking of Dynamic Objects, Diplomar thesis, KTH Royal Institut of Technology School of Electrical Engineering, Sweden, http://kth.diva-portal.org/smash/get/diva2:931627/FULLTEXT01.pdf (2016).
61. Feng, K., Zhang, G. & Cheng, B. An on-board system for detecting driver drowsiness based on multi-sensor data fusion using Dempster-Shafer theory, in 2009 International Conference on Networking, Sensing and Control, pp. 897–902 (2009).
62. Judgade, S. C. & Villortino, A. C. Grid based estimation of decision uncertainty of autonomous driving systems using belief function theory. IFAC-PapersOnLine 51, 261–266. https://doi.org/10.1016/j.ifacol.2018.07.043 (2018).
63. Lu, C., Wang, S. & Wang, X. A multi-source information fusion fault diagnosis for aviation hydraulic pump based on the new evidence similarity distance. Aerosp. Sci. Technol. 71, 392–401 (2017).
64. Yang, J., Huang, H.-Z., He, L.-P., Zhu, S.-P. & Wen, D. Risk evaluation in failure mode and effects analysis of aircraft turbine rotor blades using Dempster-Shafer evidence theory under uncertainty. Eng. Failure Anal. 18, 2084–2092. https://doi.org/10.1016/j.engfailanal.2011.06.014 (2011).
65. Fontani, M., Bianchi, T., De Rosa, A., Piva, A. & Barni, M. A Framework for decision fusion in image forensics based on Dempster-Shafer theory of evidence. IEEE Trans. Inf. Forensics Secur. 8, 593–607. https://doi.org/10.1109/TIFS.2013.2248727 (2013).
66. Chandana, S., Leung, H. & Trpkov, K. Staging of prostate cancer using automatic feature selection, sampling and Dempster-Shafer fusion. Cancer Inform. 7, 57–73 (2009).
67. Raza, M., Gondal, I., Green, D. & Coppel, R. L. Fusion of FNA-cytology and gene-expression data using Dempster-Shafer Theory of evidence to predict breast cancer tumors. Bioinformation 1, 170–175 (2006).
68. van Vliet, R. M. et al. Pooling breast cancer datasets has a synergistic effect on classification performance and improves signature stability. BMC Genom. 9, 375 (2008).
69. Rung, J. & Brazma, A. Reuse of public genome-wide gene expression data. Nat. Rev. Genet. 14, 89–99 (2013).
70. Edgar, R. & Barrett, T. NCBI GEO standards and services for microarray data. Nat. Biotechnol. 24, 1471–1472 (2006).
71. Barrett, T. & Edgar, R. Mining microarray data at NCBI’s gene expression omnibus (GEO). Methods Mol. Biol. 338, 175–190 (2006).
72. Irizarry, R. A. et al. Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res. 31, e15 (2003).
73. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy—analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 20, 307–315 (2004).
74. Stafford, P. Methods in Microarray Normalization (CRC Press, Boca Raton, 2008).
75. McCall, M. N., Jaffe, H. A. & Irizarry, R. A. IRMA ST: frozen robust multiarray analysis for Affymetrix Exon and Gene ST arrays. Bioinformatics (Oxford, England) 28, 3153–3154. https://doi.org/10.1093/bioinformatics/bts388 (2012).
76. Bolstad, B. Background and Normalization: Investigating the effects of preprocessing on gene expression estimates, http://bmbolstad.com/stuff/BAUGM.pdf (2017).
77. Kenn, M. et al. Microarray normalization revisited for reproducible breast cancer biomarkers. Biomed. Res. Int. 2020, 1363827. https://doi.org/10.1155/2020/1363827 (2020).
78. Hastie, T., Tibshirani, R. & Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction 2nd edn. (Springer, Berlin, 2009).
79. McCullagh, P. & Nelder, J. A. in Generalized Linear Models (Chapman and Hall/CRC, London, 1989).
80. Gendoo, D. M. et al. Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer.
81. Yang, J. B. & Xu, D. L. Evidential reasoning rule for evidence combination. Artif. Intell. 205, 1–29 (2013).

Acknowledgements
The authors gratefully acknowledge advice given by Georg Dorffner regarding decision theory application.

Author contributions
Conceptualization: W.S., H.K., C.S.; Methodology and software: M.K.; Formal analysis: M.K.; Data curation: M.K. and M.C.; Writing—original draft preparation: W.S.; Writing—review and editing: D.C.C., R.K., M.C. and W.S.; Supervision: W.S.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at (https://doi.org/10.1038/s41598-021-82418-7).
Correspondence and requests for materials should be addressed to W.S.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021