Fisher information.

Consider the problem of finding the best estimate of a (true) parameter ϑ from multiple measurements of experimental data, collected as elements of the data vector $\vec{y}$. As any experiment used to collect $\vec{y}$ includes error, each element of $\vec{y}$ is associated (randomly) with this experimental error. These random variations can be collected into an error vector $\vec{\epsilon}$. Then the following equation is "inevitable" description of any practical data:

$$\vec{y} = \vartheta + \vec{\epsilon} \tag{1}$$

We want to implement the best estimate of ϑ. This is accomplished by some algorithm $\hat{\theta}(\vec{y}) \rightarrow \vartheta$. Common example of such algorithm might be $\hat{\theta}(\vec{y}) \equiv \frac{1}{N}\sum_{n=1}^{N} y_n$, with y_n being the components of $\vec{y}$. The general criterion for quality of the algorithm is that it generates unbiased estimate of ϑ. Quantitatively that means the following (brackets $\langle\ \rangle$ represent mean value):

$$\langle \hat{\theta}(\vec{y}) - \vartheta \rangle = 0 \tag{2}$$

To compute the mean (2), we need probability density

$$\rho \equiv \rho(\vec{y}|\vartheta) \tag{3}$$

which describes the likelihood of measuring ϑ when $\vec{y}$ fluctuations are quantitatively described by (probabilistic) law ρ. Often ρ is Gaussian/normal distribution, but this is too restrictive, mostly relying on the conditions for central limiting theorem validity. Fisher information actually processes the information in the general form of the probability density function (not necessarily Gaussian), as is obtained from the experimental data, to do the best possible estimation of the important information, encoded in the data). The derivation of the closed algebraic form of the law describing ρ is actually the most important advantage of Fisher information based data processing.

If we assume that ρ is known, we have

$$\langle \hat{\theta}(\vec{y}) - \vartheta \rangle \equiv \int \rho\big(\hat{\theta}(\vec{y}) - \vartheta\big)\, d\vec{y} = 0 \tag{4}$$

We can now derive the definition of Fisher information from this general requirement of optimal parameter estimation. Take $\frac{\partial}{\partial\vartheta}$ of the eq. (4) above (i.e. we want to know, how this condition depends upon changes of the estimated parameter values):

$$\int \frac{\partial\rho}{\partial\vartheta}\big(\hat{\theta}(\vec{y}) - \vartheta\big)d\vec{y} + \int \rho\,\frac{\partial\big(\hat{\theta}(\vec{y}) - \vartheta\big)}{\partial\vartheta}\,d\vec{y} =$$

$$\int \frac{\partial\rho}{\partial\vartheta}\big(\hat{\theta}(\vec{y}) - \vartheta\big)d\vec{y} - \int \rho\, d\vec{y} = 0 \tag{5}$$

(Because $\frac{\partial(\hat{\theta}(\vec{y}) - \vartheta)}{\partial\vartheta} = \frac{\partial\hat{\theta}(\vec{y})}{\partial\vartheta} - \frac{\partial\vartheta}{\partial\vartheta} = 0 - 1$, as $\hat{\theta}(\vec{y})$ does not explicitly depend on ϑ). By normalization condition $\int \rho\, d\vec{y} = 1$ for probability density and using $\frac{\partial\rho}{\partial\vartheta} = \rho\frac{\partial \ln(\rho)}{\partial\vartheta}$ in (5) we have:

$\int \rho \frac{\partial \ln(\rho)}{\partial \vartheta}\left(\hat{\theta}(\vec{y}) - \vartheta\right) d\vec{y} - 1 = 0$ and finally

$$\int \rho \frac{\partial \ln(\rho)}{\partial \vartheta}\left(\hat{\theta}(\vec{y}) - \vartheta\right) d\vec{y} = 1$$

(6)

This result can be converted into the "information uncertainty principle" equation by "compartmentalization of density ρ into $\rho = \sqrt{\rho}\sqrt{\rho}$, squaring resulting form of (6) and using Schwarz inequality:

$$\left[\int \sqrt{\rho} \frac{\partial \ln(\rho)}{\partial \vartheta} \sqrt{\rho}\left(\hat{\theta}(\vec{y}) - \vartheta\right) d\vec{y}\right]^2 = 1 \leq \int \rho \left(\frac{\partial \ln(\rho)}{\partial \vartheta}\right)^2 d\vec{y} . \int \rho\left(\hat{\theta}(\vec{y}) - \vartheta\right)^2 d\vec{y}$$

(7)

with result

$$\int \rho \left(\frac{\partial \ln(\rho)}{\partial \vartheta}\right)^2 d\vec{y} . \int \rho\left(\hat{\theta}(\vec{y}) - \vartheta\right)^2 d\vec{y} \geq 1$$

(8)

In (8), we use $\frac{\partial \ln(\rho)}{\partial \vartheta} = \frac{1}{\rho}\left(\frac{\partial \rho}{\partial \vartheta}\right)$

$$\int \frac{\rho}{\rho^2}\left(\frac{\partial \rho}{\partial \vartheta}\right)^2 d\vec{y} . \int \rho\left(\hat{\theta}(\vec{y}) - \vartheta\right)^2 d\vec{y} = \int \frac{1}{\rho}\left(\frac{\partial \rho}{\partial \vartheta}\right)^2 d\vec{y} . \int \rho\left(\hat{\theta}(\vec{y}) - \vartheta\right)^2 d\vec{y} \geq 1$$

(9)

In (9), the second term is the mean squared error $\langle \varepsilon^2 \rangle$ in parameter ϑ estimation, while the first term defines Fisher information $\boldsymbol{I_F}$:

$$I_F = \int \frac{1}{\rho}\left(\frac{\partial \rho}{\partial \vartheta}\right)^2 d\vec{y}$$

(10)

The meaning of $\boldsymbol{I_F}$ can be best understood using its discrete form. We replace $\partial\vartheta$ by the (constant) difference $\Delta\vartheta = \vartheta_{n+1} - \vartheta_n$ of the estimated parameter values and integration is replaced by sum over all states of the measured system:

$$I_F = \frac{1}{\Delta\vartheta} \sum_{n=1}^{N} \frac{[\rho(\vartheta_{n+1}) - \rho(\vartheta_n)]^2}{\rho(\vartheta_n)}$$

(11)

The following algebra leads to the desired transformation of eq. (11): We first multiply (11) by $1 = \frac{\rho(\vartheta_n)}{\rho(\vartheta_n)}$ and then work out the solution:

$$I_F = \frac{1}{\Delta\vartheta} \sum_{n=1}^{N} \frac{\rho(\vartheta_n)}{\rho(\vartheta_n)} \frac{[\rho(\vartheta_{n+1}) - \rho(\vartheta_n)]^2}{\rho(\vartheta_n)} = \frac{1}{\Delta\vartheta} \sum_{n=1}^{N} \rho(\vartheta_n) \left[\frac{\rho(\vartheta_n + \Delta\vartheta)}{\rho(\vartheta_n)} - \frac{\rho(\vartheta_n)}{\rho(\vartheta_n)}\right]^2 =$$

$$= \frac{1}{\Delta\vartheta}\sum_{n=1}^{N}\rho(\vartheta_n)\left[\frac{\rho(\vartheta_n+\Delta\vartheta)}{\rho(\vartheta_n)}-1\right]^2$$

As $\Delta\vartheta$ is selected to be small, term $\left[\frac{\rho(\vartheta_n+\Delta\vartheta)}{\rho(\vartheta_n)}-1\right]^2$ can be rewritten using expansion $ln(1+x)=x-\frac{x^2}{2}\Rightarrow x^2=2[x-ln(1+x)]$. By setting $x^2=\left[\frac{\rho(\vartheta_n+\Delta\vartheta)}{\rho(\vartheta_n)}-1\right]^2$, $\boldsymbol{I_F}$ can be reformulated as:

$$I_F=\frac{1}{\Delta\vartheta}\sum_{n=1}^{N}\rho(\vartheta_n)\left[2\left(\frac{\rho(\vartheta_n+\Delta\vartheta)}{\rho(\vartheta_n)}-1\right)-ln\left(\frac{\rho(\vartheta_n+\Delta\vartheta)}{\rho(\vartheta_n)}\right)\right]=$$

$$-\frac{2}{\Delta\vartheta}\sum_{n=1}^{N}\rho(\vartheta_n)ln\left(\frac{\rho(\vartheta_n+\Delta\vartheta)}{\rho(\vartheta_n)}\right)$$

$$+\frac{2}{\Delta\vartheta}\sum_{n=1}^{N}\frac{\rho(\vartheta_n)}{\rho(\vartheta_n)}\rho(\vartheta_n+\Delta\vartheta)$$

$$-\frac{2}{\Delta\vartheta}\sum_{n=1}^{N}\rho(\vartheta_n)=-\frac{2}{\Delta\vartheta}\sum_{n=1}^{N}\rho(\vartheta_n)ln\left(\frac{\rho(\vartheta_n+\Delta\vartheta)}{\rho(\vartheta_n)}\right)+1-1=$$

$$-\frac{2}{\Delta\vartheta}\sum_{n=1}^{N}\rho(\vartheta_n)ln\left(\frac{\rho(\vartheta_n+\Delta\vartheta)}{\rho(\vartheta_n)}\right)=-\frac{2}{\Delta\vartheta^2}\sum_{n=1}^{N}\Delta\vartheta\rho(\vartheta_n)ln\left(\frac{\rho(\vartheta_n+\Delta\vartheta)}{\rho(\vartheta_n)}\right)$$

By $lim\ \Delta\vartheta\to 0$ we have

$$I_F=-\frac{2}{\Delta\vartheta^2}\int d\,\vartheta.\rho(\vartheta_n).ln\left(\frac{\rho(\vartheta_n+\Delta\vartheta)}{\rho(\vartheta_n)}\right)=-\frac{2}{\Delta\vartheta^2}KL\big(\rho(\vartheta_n),\rho(\vartheta_n+\Delta\vartheta)\big) \tag{12}$$

which is Kullback-Leibler relative entropy, describing the change of the system information when the system is in the status, characterized by parameter ϑ_n relatively to the system status, characterized by the parameter changed from that initial value to $\vartheta_n+\Delta\vartheta$.

From this relationship of the Fisher information to Kullback-Leibler relative entropy we obtained two important properties of the I_F:

- Fisher information is entropy (with all important formal properties, such as being additive etc.)
- Using the known proven properties of Kullback-Leibler relative entropy KL, we have (because of minus sign of the constant before integral in eq. (12):

$$\frac{dKL}{dt}\geq 0\ \Rightarrow\ \frac{dI_F}{dt}\leq 0 \tag{13}$$

To complete the comparison of the Fisher entropy as a function of time with all other information measures, we reiterate that for Shannon entropy *H*:

$$\frac{dH}{dt}\geq 0 \tag{14}$$

The relationship (12) between Fisher information and Kullback-Leibler "cross" entropy also best explains the "local" character of the Fisher entropy: In contrast to a global descriptor, such as Shannon entropy, which integrates the information over complete distribution of observed data (signals), the Fisher information quantifies how the information about the system's state *n* changes, when the system transitioned into state *n*+1. This is quantitatively characterized by a small increment from a *n*-th original state-specific parameter value (ϑ_n) to a new value $\vartheta_{n+1} = (\vartheta_n + \Delta\vartheta)$. This has very important consequence for the ability to derive physically, biologically and clinically relevant laws from the Fisher information. As the state-characteristic parameter increase $\Delta\vartheta$ can be selected to be very small (in the limit allowing integration actually infinitesimally small), we can use simple relationships between the functional (unknown, "hidden") and experimental (clinical) parameters, entering them into the derivation of the laws. This enables estimating the "hidden" information, such as time to disease onset, computed from the measured parameter values, such as CT-scan determined tumor mass. While in general such relationships between hidden and experimental data can be very complicated, non-linear etc., in the Fisher information processing we deal only with the small change $\Delta\vartheta$ of the processed parameter variability. In this way, it is fully justified to use simple relationships between the observed and "hidden" parameters (generally valid recipe is using just the first terms of the Taylor expansion of this complex relationship, resulting in the proportionality between the variables etc.). We thus mathematically correctly decompose a complicated relationship into piece-wise linearized series of relationships for consecutive steps of the system state development and then use calculus to generalize that discrete representation into the final law.

Another insight into the meaning of the Fisher information can be obtained by substitution of the probability density ρ by the probability amplitude for function ψ:

$$\rho(\vartheta) = \psi^2(\vartheta) \tag{15}$$

which provides simple logarithmic transformations between the two equivalent descriptors of the probability law, determining the estimate of ϑ:

$$ln\big(\rho(\vartheta)\big) = 2ln\big(\psi(\vartheta)\big)$$

and

$$ln\big(\psi(\vartheta)\big) = \frac{1}{2}ln\big(\rho(\vartheta)\big) \tag{16}$$

Substituting (15) into definition of I_F we obtain:

$$I_F = \int \frac{1}{\rho(\vartheta)}\left(\frac{\partial\rho(\vartheta)}{\partial\vartheta}\right)^2 d\vec{y} = \int \frac{1}{\psi^2(\vartheta)}\left(2\psi(\vartheta)\frac{\partial\psi(\vartheta)}{\partial\vartheta}\right)^2 d\vec{y} = \int \frac{4\psi^2(\vartheta)}{\psi^2(\vartheta)}\left(\frac{\partial\psi(\vartheta)}{\partial\vartheta}\right)^2 d\vec{y}$$
$$= 4\int \left(\frac{\partial\psi(\vartheta)}{\partial\vartheta}\right)^2 d\vec{y} \tag{17}$$

This (equivalent) form of the Fisher information shows that it is integrated gradient of the probability amplitude, governing the parameters ϑ (and thus also of the gradient of the probability density defining the relationship between measurements $\vec{y}$ and parameters ϑ).

Because of the eq. (8)

$$\int \frac{1}{\rho}\left(\frac{\partial \rho}{\partial \vartheta}\right)^2 d\vec{y} \,.\int \rho\left(\hat{\theta}(\vec{y}) - \vartheta\right)^2 d\vec{y} \geq 1$$

we derived the uncertainty relationship $I_F.\int \rho\left(\hat{\theta}(\vec{y}) - \vartheta\right)^2 d\vec{y} = \; I_F.\langle \varepsilon^2 \rangle \; \geq 1$ between the estimation error and Fisher entropy:

$$I_F.\langle \varepsilon^2 \rangle \geq 1 \tag{18}$$

which relates the Fisher information I_F to the mean squared error $\langle \varepsilon^2 \rangle$ in parameter ϑ estimation. From this relationship, we see the condition for the minimal mean squared error $\langle \varepsilon^2 \rangle$ in the ϑ determination:

$$\langle \varepsilon^2 \rangle_{min} = \frac{1}{I_F} \tag{18}$$

This identifies Fisher information as the sensitive measure of the parameter ϑ estimation.

The following important properties are derived from the fact that Fisher information is indeed entropy, so it has all the convenient properties, enabling its application to characterizing the general, complex systems:

- I_F is additive, thus if I_F^n characterizes state *n* of the studied system, the system Fisher information is

$$I_F = \sum_n I_F^n \tag{19}$$

- For multidimensional systems, using the addictiveness of I_F^n, the definition (17) becomes

$$I_F = 4\int \sum_n \left(\vec{\nabla}\psi_n . \vec{\nabla}\psi_n\right) d\vec{x} \tag{20}$$

  where $\vec{x} = [x_0, \ldots, x_k]$ and $\nabla_k = \frac{\partial}{\partial x_k}$.

- For mixed systems, the Fisher information obeys triangular inequality, so it is important distance measure, characterizing differences in composition of these systems:

$$aI_F(\psi_1) + bI_F(\psi_2) \geq I_F(a\psi_1 + b\psi_2) \tag{21}$$

  As a consequence,

$$I_F\left(\sum_{n=1}^{N} A_n\psi_n\right) \leq \sum_{n=1}^{N} A_n I_F(\psi_n) \tag{22}$$

The following considerations of the **E**xtreme **P**hysical **I**nformation (EPI) principle allowed to convert the Fisher information into law-generating tool. The EPI approach logically reverses the understanding of the data measurement process: instead treating it as collecting the random output from the (physical) phenomenon or effect, EPI allows reverse tracking of the function-related information from output to input, from the data to the underlying effect. It uses knowledge of the information flow in the measurement process to derive the mathematical form of the physical effect that gives the output measurements. It assumes, though, that the effect "is out there" and the result is a mathematical expression, quantitatively characterizing that effect. It will not lead to formulating new effects without experimental evidence for their existence.

As the analysis of algebras of complex systems indicates, the information representing the actual phenomena can be identified by the invariance (or symmetries) of the corresponding representations. This proves the active role of invariants in the deriving the physical laws. It again is a reverse of conventional deductive approaches, where the law is "created" by thought experiment and its relevance is then confirmed by showing its (required) invariances.

This qualitative statement has simple quantitative formulation. To be able to express the EPI, the information functional, captured by the Fisher entropy is partitioned into two components – internal (***I***) and external (***J***). The internal Fisher information is related to the physical, biological, systemic processes, which generate non-random variation(s) of the measured parameter. The external Fisher information component captures all aspects of the parameter measurement process. In general, the physically (biologically, clinically) relevant information flow goes from the fundamental internal processes to the experimental results, formally $I \to J$. The balance between the internal processes and how they can be identified in the measured data is captured by the following equation:

$$J = \kappa I \tag{23}$$

where κ quantifies the extent of information lost, $\kappa \leq 1$ (e.g. by the incomplete conversion of the information during the experiments) or "gained" in the interacting systems, where given measurement is carrying contributions from several (or all) other system states, $\kappa > 1$.

The functionals ***I*** and ***J*** are defined by the following integral equations:

$$J \equiv 4\int j(t)dt = 4\int \left(\frac{\partial\psi}{\partial t}\right)^2 dt \Rightarrow j(t) = \dot{\psi}^2 \tag{24}$$

$$I \equiv 4\int i(\psi, t)dt \tag{25}$$

Essential role of ***I*** and ***J*** is in fact, that requirement (23) represents physically optimal situation, when experimental data collection is done with no (or minimal) loss of the information about the essential (non-random) internal processes. In formal sense, requirement

$$J - \kappa I = 0 \tag{26}$$

is the quantitative formulation of the Extreme Physical Information (EPI) principle, allowing implementation of the whole Fisher information based calculus. As the information optimality criterion is expressed in terms of integral equations (24) and (25), this condition has to be met through extreme forms of ***I*** and ***J***. This second requirement leads first to the general Euler-Lagrange equation, which is then adapted to process ***I*** and ***J***, as is derived next.

Define integral function ***K***, whose extreme we need to find

$$K = \int_a^b dx \mathcal{L}\left(x, \psi(x), \dot{\psi}(x)\right) = extreme \tag{27}$$

To introduce needed "variability" into ***K***, consider perturbed density amplitude

$$\psi_\varepsilon(x, \varepsilon) = \psi(x) + \varepsilon\eta(x) \tag{28}$$

We require that at the boundaries *a*,*b* of the measurement interval, the $\psi_\varepsilon(x, \varepsilon) = \psi(x)$. Thus, perturbing function $\eta(a) = \eta(b) = 0$. With these conditions, we can look for extreme of (ε) :

$$K(\varepsilon) = \int_a^b dx \mathcal{L}\left(x, \psi_\varepsilon(x, \varepsilon), \dot{\psi}_\varepsilon(x, \varepsilon)\right) \tag{29}$$

which requires $\frac{\partial K}{\partial \varepsilon} = 0$ at $\varepsilon = 0$. Computing this derivation of (29) gives

$$\begin{aligned}\frac{\partial K}{\partial \varepsilon} &= \int_a^b dx \left[\frac{\partial \mathcal{L}}{\partial \psi_\varepsilon}\frac{\partial \psi_\varepsilon}{\partial \varepsilon} + \frac{\partial \mathcal{L}}{\partial \dot{\psi}_\varepsilon}\frac{\partial \dot{\psi}_\varepsilon}{\partial \varepsilon}\right] = \int_a^b dx \left[\frac{\partial \mathcal{L}}{\partial \psi_\varepsilon}\frac{\partial \psi_\varepsilon}{\partial \varepsilon}\right] + \int_a^b dx \left[\frac{\partial \mathcal{L}}{\partial \dot{\psi}_\varepsilon}\frac{\partial \dot{\psi}_\varepsilon}{\partial \varepsilon}\right] \\ &= \int_a^b dx \left[\frac{\partial \mathcal{L}}{\partial \psi_\varepsilon}\frac{\partial \psi_\varepsilon}{\partial \varepsilon}\right] + \int_a^b dx \left[\frac{\partial \mathcal{L}}{\partial \dot{\psi}_\varepsilon}\frac{\partial^2 \psi_\varepsilon}{\partial x \partial \varepsilon}\right]\end{aligned}$$

Integrating *per partes* the second integral results in

$$\frac{\partial \mathcal{L}}{\partial \dot{\psi}_\varepsilon}\frac{\partial \psi_\varepsilon}{\partial \varepsilon}|_a^b - \int_a^b dx \frac{\partial \psi_\varepsilon}{\partial \varepsilon}\frac{d}{dx}\left(\frac{\partial \mathcal{L}}{\partial \dot{\psi}_\varepsilon}\right)$$

as $\frac{\partial \psi_\varepsilon}{\partial \varepsilon}|_a^b = 0$, we have

$$\begin{aligned}\frac{\partial K}{\partial \varepsilon} &= \int_a^b dx \left[\frac{\partial \mathcal{L}}{\partial \psi_\varepsilon}\frac{\partial \psi_\varepsilon}{\partial \varepsilon} - \int_a^b dx \frac{\partial \psi_\varepsilon}{\partial \varepsilon}\frac{d}{dx}\left(\frac{\partial \mathcal{L}}{\partial \dot{\psi}_\varepsilon}\right)\right] = \int_a^b dx \frac{\partial \psi_\varepsilon}{\partial \varepsilon}\left[\frac{\partial \mathcal{L}}{\partial \psi_\varepsilon} - \frac{d}{dx}\left(\frac{\partial \mathcal{L}}{\partial \dot{\psi}_\varepsilon}\right)\right] \\ &= \int_a^b dx\ \eta(x)\left[\frac{\partial \mathcal{L}}{\partial \psi_\varepsilon} - \frac{d}{dx}\left(\frac{\partial \mathcal{L}}{\partial \dot{\psi}_\varepsilon}\right)\right] = 0\end{aligned}$$

As $\eta(x)$ is arbitrary, this integral can be zero only when

$$\frac{\partial \mathcal{L}}{\partial \psi_\varepsilon} = \frac{d}{dx}\left(\frac{\partial \mathcal{L}}{\partial \dot{\psi}_\varepsilon}\right) \tag{30}$$

Eq. (30) is the Euler-Lagrange equation. For processing Fisher information, we set $\mathcal{L} = j - i$:

$$\frac{d}{dt}\left(\frac{\partial(j-i)}{\partial\dot{\psi}}\right) = \frac{\partial(j-i)}{\partial\psi} \quad and \quad j - \kappa i = \dot{\psi}^2 - \kappa i(\psi, t) = 0 \tag{31}$$

where $i = i(\psi, t)$, $j = \dot{\psi}^2$ and solve simultaneously:

$$\frac{d}{dt}\left(\frac{\partial\left(\dot{\psi}^2 - i(\psi,t)\right)}{\partial\dot{\psi}}\right) = \frac{\partial\left(\dot{\psi}^2 - i(\psi,t)\right)}{\partial\psi}, \frac{\partial\dot{\psi}^2}{\partial\psi} - \frac{\partial i(\psi,t)}{\partial\psi} = -\frac{\partial i(\psi,t)}{\partial\psi}$$

$$-\frac{\partial i(\psi,t)}{\partial\psi} = \frac{d}{dt}\left(\frac{\partial\left(\dot{\psi}^2 - i(\psi,t)\right)}{\partial\dot{\psi}}\right) = \frac{d}{dt}\left(\frac{\partial\dot{\psi}^2}{\partial\dot{\psi}} - \frac{\partial i(\psi,t)}{\partial\dot{\psi}}\right) = \frac{d}{dt}(2\dot{\psi}) = 2\frac{d}{dt}\left(\frac{d\psi}{dt}\right) = 2\frac{d^2\psi}{dt^2}$$

$$2\frac{d^2\psi}{dt^2} = -\frac{\partial i(\psi,t)}{\partial\psi} \tag{32}$$

Next we use $\dot{\psi}^2 = \kappa i(\psi, t)$ and differentiate both sides by $^d/_{dt}$ and use (32) in the result:

$$2\dot{\psi}\frac{d^2\psi}{dt^2} = \kappa\left[\frac{\partial i}{\partial t} + \frac{\partial i}{\partial\psi}\frac{\partial\psi}{\partial t}\right] \Rightarrow -\dot{\psi}\frac{\partial i}{\partial\psi} = \kappa\left[\frac{\partial i}{\partial t} + \frac{\partial i}{\partial\psi}\dot{\psi}\right] \tag{33}$$

From $\dot{\psi}^2 - \kappa i(\psi, t) = 0$ we have $\dot{\psi} = \sqrt{\kappa}\sqrt{i(\psi,t)}$ and (33) becomes

$$-\sqrt{\kappa}\sqrt{i}\,\frac{\partial i}{\partial\psi} = \kappa\left[\frac{\partial i}{\partial t} + \frac{\partial i}{\partial\psi}\sqrt{\kappa}\sqrt{i}\right]$$

$$\frac{-\sqrt{\kappa}}{\sqrt{\kappa}}\sqrt{i}\,\frac{\partial i}{\partial\psi} = \frac{\kappa}{\sqrt{\kappa}}\frac{\partial i}{\partial t} + \frac{\kappa\sqrt{\kappa}}{\sqrt{\kappa}}\sqrt{i}\,\frac{\partial i}{\partial\psi}$$

$$-\sqrt{i}\,\frac{\partial i}{\partial\psi} = \sqrt{\kappa}\frac{\partial i}{\partial t} + \kappa\sqrt{i}\,\frac{\partial i}{\partial\psi}$$

$$-\sqrt{i}\,\frac{\partial i}{\partial\psi} - \kappa\sqrt{i}\,\frac{\partial i}{\partial\psi} = \sqrt{\kappa}\frac{\partial i}{\partial t}$$

$$-\sqrt{i}\,\frac{\partial i}{\partial\psi}(1+\kappa) = \sqrt{\kappa}\frac{\partial i}{\partial t}$$

$$\sqrt{\kappa}\frac{\partial i}{\partial t} + \sqrt{i}\,\frac{\partial i}{\partial\psi}(1+\kappa) = 0 \tag{34}$$

Differential equation (34) can be solved by separation of variables, using

$$i = i_\psi(\psi) i_t(t) \tag{35}$$

Substituting (35) into (34) gives

$$(1+\kappa)\sqrt{i_\psi i_t}\frac{\partial i_\psi}{\partial \psi}i_t + \sqrt{\kappa}\frac{\partial i_t}{\partial t}i_\psi = 0$$

$$(1+\kappa)i_\psi^{\frac{1}{2}} i_t^{\frac{3}{2}}\frac{\partial i_\psi}{\partial \psi} + \sqrt{\kappa}\frac{\partial i_t}{\partial t}i_\psi = 0$$

multiplying by $\frac{1}{i_\psi}.\frac{1}{i_t^{3/2}}$ gives

$$(1+\kappa)\frac{1}{\sqrt{i_\psi}}\frac{\partial i_\psi}{\partial \psi} + \sqrt{\kappa}\frac{1}{\sqrt{i_t^3}}\frac{\partial i_t}{\partial t} = 0 \tag{36}$$

The sum of the two terms on the left hand side of (36) can be equal to zero only if the both terms are equal to the same constant with opposite signs:

$$(1+\kappa)\frac{1}{\sqrt{i_\psi}}\frac{\partial i_\psi}{\partial \psi} = A, \qquad \sqrt{\kappa}\frac{1}{\sqrt{i_t^3}}\frac{\partial i_t}{\partial t} = -A \tag{37}$$

Equations (37) can be integrated

$$\int \frac{1}{\sqrt{i_\psi}}\, di_\psi = \frac{A}{(1+\kappa)}\int d\psi\,, \qquad \frac{1}{\sqrt{i_t^3}}\int di_t = -\frac{A}{\sqrt{\kappa}}\int dt$$

$$\sqrt{i_\psi} - \frac{1}{2}\frac{A\psi}{(1+\kappa)} + B = 0$$

Squaring

$$i_\psi - \frac{1}{4}\left(\frac{A\psi}{(1+\kappa)} + B\right)^2 = 0 \qquad \Rightarrow \quad i_\psi = \frac{1}{4}\left(\frac{A\psi}{(1+\kappa)} + B\right)^2 \tag{37}$$

For i_t we get similarly

$$i_t = 4\left(\frac{At}{\sqrt{\kappa}} + C\right)^{-2} \tag{38}$$

Combining (37) and (38) into (35) results in

$$i(\psi, t) = i_\psi . i_t = \frac{4}{4}\left(\frac{\frac{A\psi}{(1+\kappa)} + B}{\frac{At}{\sqrt{\kappa}} + C}\right)^2$$

(39)

Because $\frac{d\psi}{dt} = \sqrt{i(\psi,t)}$ and thus $d\psi = \sqrt{i(\psi,t)}dt$, (39) becomes

$$\frac{d\psi}{\frac{A\psi}{(1+\kappa)} + B} = \frac{dt}{\frac{At}{\sqrt{\kappa}} + C}$$

which is integrated to obtain ψ:

$$\psi = \left(\frac{(1+\kappa)}{A}\right)\left(\frac{At}{\kappa} + D\right)^{\left(\frac{\kappa}{1+\kappa}\right)} - (1+\kappa)\frac{B}{A}$$

(40)

Some constants in the general solution (40) can be found by boundary conditions: at *t*=0, *B*=0, *D*=0, giving finally

$$\psi(t) = \left(\frac{1+\kappa}{A}\right)\left(\frac{A}{\kappa}\right)t^{\left(\frac{\kappa}{1+\kappa}\right)} = Et^{\alpha}$$

(41)

Time-dependent processes (such as tumor growth, plaque deposit, drug transport etc.) encode into the Fisher entropy the information about the functional (intrinsic) and assay-dependent (extrinsic) factors, influencing the probability densities of the dependent clinical variable values. Fisher information allows quantitative characterization and extraction of this unique information from the experimental probability density estimates. What follows is the outline of the underlying mathematics.

In a growing system, there are *N* types of entities/states (patients, cells, etc.), which provide the information into the Fisher entropy-based data processing. The set of these states is finite, with cardinality *M*. This is expressed as follows:

$$M = \sum_{n=1}^{N} m_n$$

(42)

Probability density ρ_n of values, related to one state *n* is defined as the frequency of observing a value, characteristic for that state *n*:

$$\rho_n = \frac{m_n}{M} \tag{43}$$

The complete system description uses the probability density vector $\vec{\rho} = [\rho_1, \dots, \rho_N]$. The probability density has the following properties – for growth and transport processes, $\rho_n = \rho_n(t)$ i.e. it is a function of time. ρ_n is also normalized, from which the following property of its time derivatives (gradients) follows:

$$\sum_{n=1}^{N} \rho_n = 1 \Rightarrow \sum_{n=1}^{1} \frac{d^i \rho_n}{dt^i} = 0$$

(44)

Analysis of the components of the overall probability density $\vec{\rho}$ provides the fundamental description of its dynamics using the following *N* equations:

$$\frac{d\rho_n}{dt} = \dot{\rho}_n = \rho_n(g_n + d_n) \tag{45}$$

where g_n is creation coefficient and d_n is depletion coefficient, characterizing quantitatively processes, in which the system states, generating characteristic parameter values are emerging (for example a patient's disease progressed from 'moderate' to 'serious' status, accompanied by the increase of the clinical status marker values above certain threshold(s)). In simple terms, this relationship indicates, that the rate, with which the probability of assaying certain clinical state is changing, is proportional to the probability density ρ_n of experimental finding that result. The extent of this proportionality between the density dynamics and density itself is dependent upon equilibrium between the state creation and annihilation processes in the studied patient. Essential for the application of Fisher information paradigm in the Personalized Network Medicine is the fact that the state creation and depletion processes, described quantitatively by functions g_n and d_n, depend on the interactions between all states of the concrete individual patient. To quantify this important statement, we realize that state creation and depletion parameters in the equation (45) are dependent upon all other states of the system:

$$g_n = \sum_{k=1}^{K \le N} g_{nk}\, \rho_k$$

with $$g_{nk} \equiv g_{nk}(\vec{\rho}, t), \qquad d_n \equiv d_n(\vec{\rho}, t) \tag{46}$$

Computing $\rho_n = \psi_n^2$ from tumor growing law gives $\rho_n = E^2 t^{2\alpha} = \tilde{E} t^{\gamma}$. Differentiation by time allows checking if the solution (41), applied to a concrete problem of tumor growth, satisfies the general growth relationship (45) :

$$\dot{\rho}_n = \frac{d\tilde{E}t^{\gamma}}{dt} = \tilde{E}\gamma t^{(\gamma-1)}$$

This equals to (45) with $(g_n + d_n) = \frac{\gamma}{t}$:

$$\tilde{E}\gamma t^{(\gamma-1)} = \tilde{E}t^{\gamma}\left(\frac{\gamma}{t}\right) \tag{47}$$

By identifying $g(\rho_n, t) = \left(\frac{\kappa}{1+\kappa}\right)\left(\frac{1}{t}\right)$, we see that the first term indicates the extent of interacting (biological) states, contributing to the tumor growth and the second, time dependent term, indicates that with increasing time the creation rate of the tumor slows down (in agreement with the known facts: Studies have shown that tumor growth rate may decline with time (Hart, Shochat et al. 1998;Bajzer 1999; Afenya and Calderon 2000), which results in non-exponential growth model of tumors. Growth deceleration has been observed in animal models (Wennerberg, Willen et al.1988), for solid tumors in clinical studies (Spratt, von Fournier et al. 1993; Spratt, Meyer et al.1996), and in leukemia (Afenya and Calderon 2000). Growth deceleration is attributed to several factors, including prolonged cell cycle, reduced growth fraction, decreased availability of oxygen (Pavelic, Porter et al. 1978), decreased cell

proliferation rate with increased cell loss rate (Bassukas and Maurer-Schultze 1987), tumor-related systemic factors (DeWys 1972), and allometric growth control (Prehn 1991)).

In eq. (47), ***E*** and γ are constants to be determined from study data. ***E*** is the cancer-type specific "amplitude", summarizing all non-tumor mass related tumorigenesis processes and $\gamma = \left(\frac{\kappa}{1+\kappa}\right)$ is the rate constant, describing the probability gradient rate of finding certain tumor mass in the patient's cohort at baseline (when CT scan of the tumor was taken). Note (for interpretation purposes) that γ is solely determined by κ, the parameter, which is proportional to the number of all INTERACTING intercellular biological processes, influencing the tumor growth.

Still, both sides of (47) are (continuous and general) functions of time. To personalize this general result, we therefore need to express explicitly the change of $\psi(t)$ and $E.t^{\gamma}$ with time and integrate the resulting formulae up to the time, when the patient's tumor mass was observed in clinic. To facilitate direct comparison with tumor mass histogram, we use $\psi^2 = \rho(Tmass, t)$:

$$\int_{t=0}^{t=t_{baseline}} \rho(t) = E^2 \int_{t=0}^{t=t_{baseline}} t^{2\gamma} dt \tag{48}$$

Carrying the integration, we obtain the following functions of the (personal) upper integration limit $t_{baseline}$:

$$\int_{t=0}^{t=t_{baseline}} \rho(t)dt = P_{intg} \tag{49}$$

$$E^2 \int_{t=0}^{t=t_{baseline}} t^{2\gamma} dt = \frac{E^2\left(t_{onset}^{(2\gamma+1)} - 1\right)}{(2\gamma+1)} = P_{intg} \tag{50}$$

The personal parameter P_{intg} is obtained by integrating the (normalized) histogram of the study tumor masses, up to the value found for a patient. We can then solve the last equation for $t_{baseline}$, obtaining the final formula (Ω_o is cancer-type specific constant):

$$t_{baseline} = \Omega_o e^{\left(\frac{ln\left[\frac{2\gamma.P_{intg}+P_{intg}+E^2}{E^2}\right]}{(2\gamma+1)}\right)} \tag{51}$$

The first step in applying these theoretical result to actual data is to verify, that the actual ***Tmass*** histogram is compatible with the "power law" (47). By considering the ***Tmass*** proportional to disease duration

(which is equivalent to taking the first terms of the Taylor expansion of that relationship), we have $\rho(Tmass) = \psi^2(Tmass) = \tilde{E}^2.Tmass^{2\gamma}$, and taking the logarithms we have

$$ln\big(\rho(Tmass)\big) = ln\big(\tilde{E}^2\big) + 2\gamma.ln(Tmass) \quad (52)$$

indicating that if the (normalized) tumor mass histogram is presented in the log-log representation, the logarithm of $\rho(Tmass)$ should be linear function of logarithm of ***Tmass***, with $ln\big(\tilde{E}^2\big)$ being equal to intercept and 2γ to the slope of that relationship.

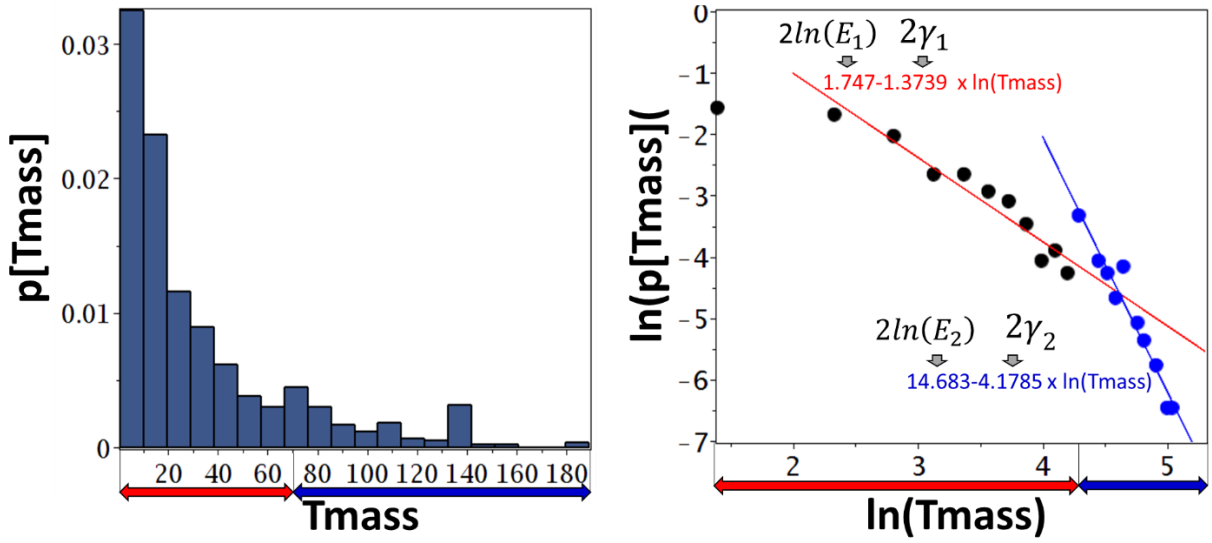
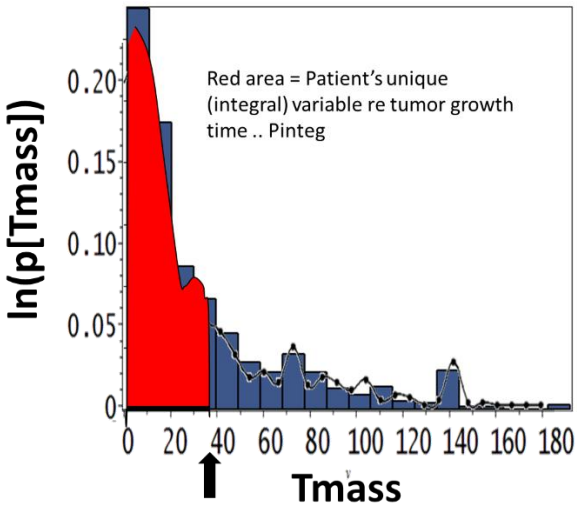
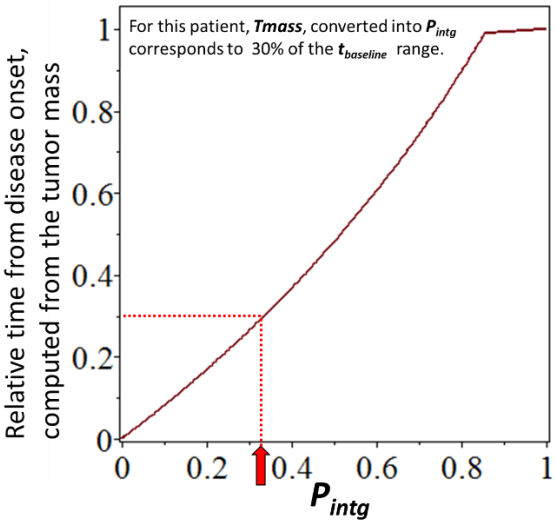


Fig. 1a) shows the original histogram for cohort of 641 HCC patients (Pittsburgh cohort, details published in refs xxx). **Fig. 1b)** shows that histogram converted into log-log scales. It is observed that these clinical data are reflecting two tumor-growth processes. The first linear relationship describes the Fisher information related processes for tumors with masses less or equal to ***Tmass***=70, the second linear relationship holds for tumors with masses larger than 70. The least-squared fits by the two linear relationships are shown, with the values of the parameters, needed to compute $t_{baseline}$, indicated.

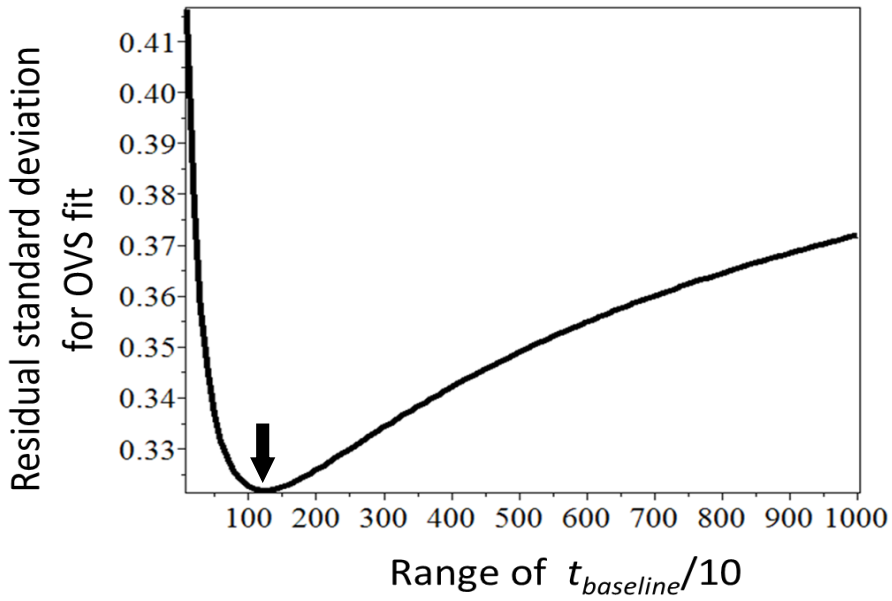
Fig. 2a) shows the use of eq. (49) to compute the P_{intg} for one patient from the histogram of observed tumor masses. **Fig. 2b)** shows the plot of $t_{baseline}$ as the function of P_{intg}, computed from the respective tumor masses, considering explicitly the presence of the two rates of the probability density gradients in eq. (51). Note that Fisher information processing indicates, that hepatocellular tumor growth is heterogeneous, but with just two significant types of growth processes. One type describes the smaller tumors, ***Tmass*** below 70, the other the tumor with masses above 70. By using the estimated values of γ_1 and γ_2 and definition $\gamma_i = \left(\frac{\kappa_i}{1+\kappa_i}\right)$, we find that $\frac{\kappa_1}{\kappa_2} = 2.8$, indicating that earlier stages of the tumor growth

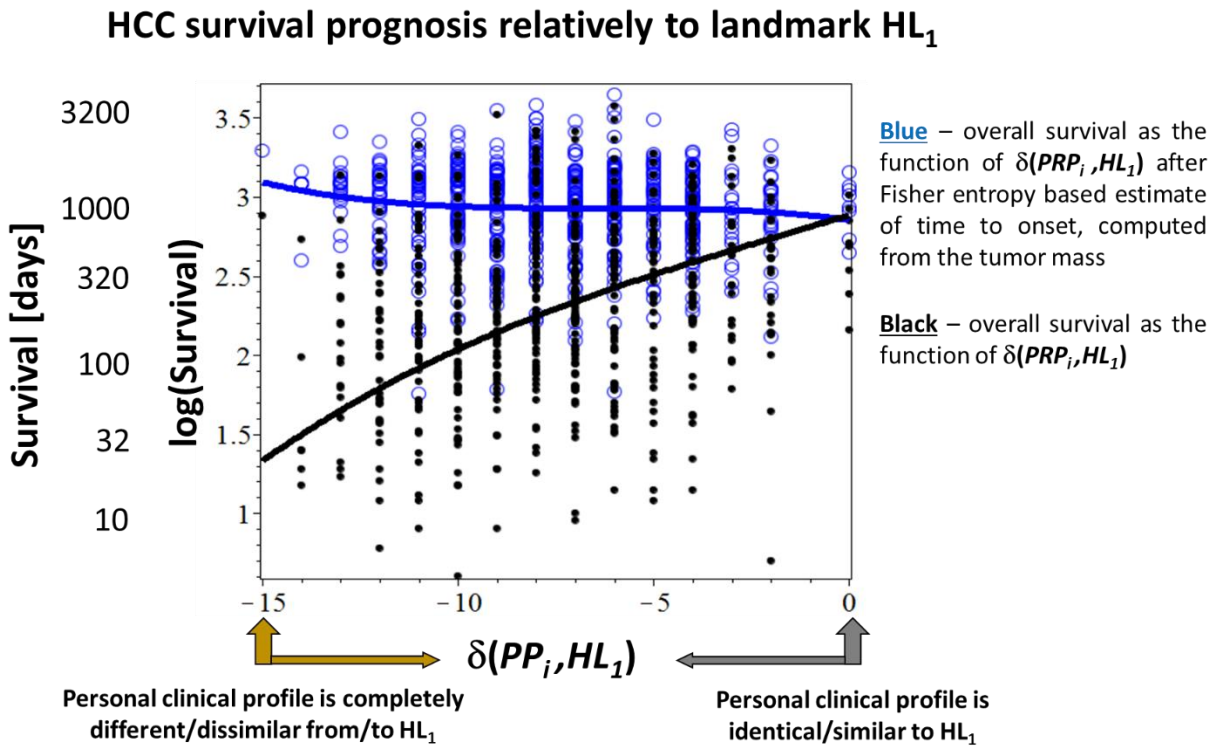




are more sensitive to the overall clinical context of the patient (e.g. in terms of micro- and macro-environmental factors, ref xx). Our Fisher entropy-based analysis thus showed that growth of these smaller tumors, which constituted about 80% of all tumors found in this US/Pittsburgh cohort, involved on average ~3 times more interacting cellular processes than were observable for multiple, very large tumors with total masses above 70, observed for remaining 20% of screened patients.

The last step of the converting of the ***Tmass*** data into $t_{baseline}$ through eq. (2) is finding the cancer-type specific constant, Ω_o, which will convert the relative time units of the exponential term of the eq. (2) to actual days. We determine Ω_o as a constant that will reproduce best the survival prognosis computed from the coherence baseline descriptors δHL_1 of our patients. This was done by the following systematic procedure: we varied Ω_o value from 0 to 10 000 days with 10 day increment. For each of these values of Ω_o, we corrected the patient's survival by the value $\boldsymbol{T_{baseline}}=\Omega_o . t_{baseline}$: $\boldsymbol{OVS_C}=\boldsymbol{OVS}-\boldsymbol{T_{baseline}}$ and fitted the resulting set of 641 $\boldsymbol{OVS_C}$ values by $\log(\boldsymbol{OVS_C}) = \boldsymbol{\beta_3} \times \delta HL_1^3 + \boldsymbol{\beta_2} \times \delta HL_1^2 + \boldsymbol{\beta_1} \times \delta HL_1 + \boldsymbol{q}$. **Fig. 3** shows the residual standard deviations of the fits $\boldsymbol{OVS_C} - \delta HL_1$, which used those systematically varied values of Ω_o. The Ω_o= 1200 days, representing the minimum residual range between the actual and prognosed $\boldsymbol{OVS_C}$ values is the value of this constant, characterizing the hepatocellular cancer.





Main result is that within the 15 clinical coherence categories, the shortest OVS are due to largest $t_{baseline}$, corresponding (by the corresponding law) to the largest tumors. What remains after the $t_{baseline}$ correction are coherence context components of the OVS, that exhibit good functional relationship (good fit) and have significantly narrower range of differences between the predicted and actual OVS.