July 2-5, 2017
London, United Kingdom

Association for Computing Machinery

Advancing Computing as a Science & Profession

DH'17

Proceedings of the 2017 International Conference on Digital Health

In cooperation with:

ACM SIGKDD & ICPS

Supported by:

UCL Institute for Risk & Disaster Reduction, UCL Grand Challenges, Frontiers in Digital Health, International Society for Telemedicine & eHealth, Mad*Pow, Digital Catapult, UK Science & Innovation Network, UCL Global Engagement Office and Enterprise Europe Network

DH 2017 Conference Organization

General and Scientific Chair: Patty Kostkova *(University College London)*

Programme Co-Chairs: Floriana Grasso *(University of Liverpool)*
Carlos Castillo *(Eurecat)*
Yelena Mejova *(QCRI)*
Arnold Bosman *(Transmissible)*

Local and Comms Chair: Patrizia Isabelle Duda *(IRDR, UCL; WADEM)*

Poster and Demo Chair: Andreea Molnar *(University of Portsmouth)*

Web and Proceedings Chair: Andrew Roxburgh *(University of Liverpool)*

PhD Track Chair: Yelena Mejova *(QCRI)*

Business Chair: Sarah Schubert *(IRDR, University College London)*

Coordinator: Emma Defoe *(University College London)*

Senior Programme Committee: Philip Abdelmalik (*Public Health Agency of Canada*)
Bruria Adini (*Tel Aviv University*)
Eyhab Al-Masri (*University of Waterloo*)
Alain April (*École de Technologie Supérieure*)
Jakob E. Bardram (*Technical University of Denmark*)
Isaac Bogoh (*University of Toronto*)
Pamela Briggs (*Northumbria University*)
Nigel Collier (*University of Cambridge*)
Simon de Lusignan (*University of Surrey*)
Dustin Ditommaso (*Mad*Pow*)
Richard Dobson (*King's College London and UCL*)
Claudio Eccher (*FBK-IRST*)
Baltasar Fernández Manjón (*Universidad Complutense de Madrid*)
Pietro Ghezzi (*Brighton & Sussex Medical School*)
Hamed Haddadi (*Queen Mary University of London*)
David Harper (*Chatham House*)
Dionisio Herrera (*TEPHINET*)
Jeroen Jansen (*Evidence Aid*)
Ilan Kelman (*University College London*)
Aileen Kitching *(EPIET Alumni Network (EAN))*
Alex Matic (*Telefonica Research*)
Eric Meyer (*University of Oxford*)
Nuria Oliver (*DataPopAlliance*)
Hans C. Ossebaard (*University of Twente/Health care Quality Institute*)
Daniela Paolotti (*ISI Foundation*)
Michal Rosen-Zvi (*IBM*)
Alberto Sanna (*Scientific Institute San Raffaele*)
Michele Tizzoni (*ISI Foundation*)
Erik van der Goot (*European Commission's Joint Research Centre*)

Programme Committee:

Harshavardhan Achrekar *(University of Massachusetts Lowell)*
Akram Alomainy *(Queen Mary University of London)*
Nadia Berthouze *(University College London)*
Olivier Bodenreider *(US National Library of Medicine)*
Celia Boyer *(Health On the Net Foundation)*
Miroslav Bursa *(Czech Technical University in Prague)*
Elena Cardillo *(Institute for Informatics and Telematics, Italian National Council of Research)*
Robin Cohen *(U Waterloo)*
Mike Conway *(University of Utah)*
Olivier Corby *(INRIA)*
Ulises Cortés *(Universitat Politècnica de Catalunya / Barcelona Supercomputing Center)*
Jean Costa *(Cornell University)*
Aaron Crandall *(Washington State University)*
Aron Culotta *(Illinois Institute of Technology)*
Mary Czerwinski *(Microsoft Research)*
Karen Day *(University of Auckland)*
Berardina Nadja De Carolis *(dipartimento di informatica universita' di bari)*
Gayo Diallo *(ISPED & LABRI, University of Bordeaux)*
Jim Duggan *(National University of Ireland, Galway)*
Michael Edelstein *(Chatham House)*
Luis Fernandez Luque *(Qatar Computing Research Institute)*
Edward Fottrell *(UCL Institute for Global Health)*
Ana Franco *(Instituto Gulbenkian de Ciência)*
Reva Freedman *(Northern Illinois University)*
Amira Ghenai *(University of Waterloo)*
Christophe Giraud-Carrier *(Brigham Young University)*
Natalia Grabar *(STL CNRS Université Lille 3)*
Cathal Gurrin *(Dublin City University)*
Jaap Ham *(Eindhoven University of Technology)*
Asha Herten-Crabb *(Chatham House)*
Helmut Horacek *(Saarland University)*
Muhammad Imran *(Qatar Computing Research Institute)*
Simon Kamronn *(Technical University of Denmark)*
Matthew Kay *(University of Washington)*
Jeonggil Ko *(Ajou University)*
Spyros Kotoulas *(IBM Research)*
Vasileios Lampos *(University College London)*
Toby Lasserson *(Cochrane)*
Marilyn Lennon *(University of Glasgow)*
Lenka Lhotska *(Czech Technical University in Prague, Dept. of Cybernetics)*
Nut Limsopatham *(Univeristy of Cambridge)*
Peter Lucas *(Radboud University Nijmegen)*
Jochen Meyer *(OFFIS Institute for Information Technology)*
Francesco Miele *(Bruno Kessler Foundation - Center for Communication and Information Technology)*
Antonio Moreno *(URV)*

Programme Committee (continued):

Bobak Mortazavi *(Texas A&M Center for Remote Health Technologies and Systems)*
Nikki Newhouse *(University College London)*
Rita Orji *(University of Saskatchewan)*
Michal Ozery-Flato *(IBM Haifa Research lab)*
Heather Patrick *(Carrot Sense)*
Pasi Penttinen *(ECDC)*
Ming-Zher Poh *(Cardiio, Inc.)*
Matthew Purver *(Queen Mary University of London)*
Luz Rello *(Carnegie Mellon University)*
David Riaño *(Universitat Rovira i Virgili)*
Britta Ricker *(University of Washington Tacoma)*
Sara Rubinelli *(University of Lugano)*
Carlos Luis Sánchez Bocanegra *(University of Seville)*
Melina Santos *(Ministry of Health of Brazil)*
Heiko Schuldt (University of Basel)
Uri Shalit *(New York University)*
Tomer Simon *(Ready.org.il)*
Steven Simske *(Hewlett-Packard Labs)*
Aneesha Singh *(UCL Interaction Centre)*
Kirsten Smith *(University of Southampton)*
Katarzyna Stawarz *(University of Bristol)*
Charlotte Tang *(University of Michigan Flint)*
Adel Taweel
Matthew E. Taylor *(Washington State University)*
Ielka Van Der Sluis *(University of Groningen)*
Kathleen Vancleef *(Newcastle University)*
Krishna K. Venkatasubramanian *(Worcester Polytechnic Institute)*
Nervo Verdezoto *(University of Leicester)*
Silja-Riin Voolma *(University of Cambridge)*
Katarzyna Wac *(University of Geneva)*
Jessica Wardlaw *(University of Nottingham)*
Ingmar Weber *(Qatar Computing Research Institute)*
Sarah Wiseman *(Goldsmiths, University Of London)*
Yeliz Yesilada *(University of Manchester)*
Qian Zhang *(Northeastern University)*

Additional Reviewers:

Alexandra Balahur *(European Commision's Joint Research Centre)*
Jo Gibbs *(UCL Interaction Centre)*
Sakib Jalil *(University of Saskatchewan)*
Laura König *(Universität Konstanz)*
Richard Lomotey *(University of Saskatchewan)*
Michaela Reisinger *(Austrian Institute of Technology)*
Beatrix Zechmann *(Austrian Institute of Technology)*
Simo Zhang *(Northeastern University)*

Table of Contents

Session: Study Methodologies

Session: Social Media

Session: Health Systems & Tools

PhD Session

On Consolidated Predictive Model of the Natural History of Breast Cancer: Primary Tumor and Secondary Metastases in Patients with Lymph Nodes Metastases

Ella Y. Tyuryumina
National Research University Higher School of Economics
Moscow, Russia
eyatyuryumina@gmail.com

Alexey A. Neznanov
National Research University Higher School of Economics
Moscow, Russia
aneznanov@hse.ru

ABSTRACT

This paper is devoted to mathematical modelling of the progression and stages of breast cancer. The âĂIJ**Co**nsolidated mathematical growth **M**odel of primary tumor (PT) and secondary distant metastases (MTS) in patients with lymph nodes MTS (Stage III)" (CoM-III) is proposed as a new research tool. The CoM-III rests on an exponential tumor growth model and consists of a system of determinate nonlinear and linear equations. The CoM-III describes correctly primary tumor growth (parameter T) and distant metastases growth (parameter M, parameter N). The CoM-III model and predictive software: a) detect different growth periods of primary tumor and distant metastases in patients with lymph nodes MTS; b) make forecast of the period of the distant metastases appearance in patients with lymph nodes MTS; c) have higher average prediction accuracy than the other tools; d) can improve forecasts on survival of breast cancer and facilitate optimisation of diagnostic tests. The CoM-III enables us, for the first time, to predict the *whole natural history* of PT and secondary distant MTS growth of patients with/without lymph nodes MTS on each stage relying only on PT sizes.

CCS CONCEPTS

• **Applied computing** → **Consumer health**; **Health informatics**; • **Mathematics of computing** → *Solvers*;

KEYWORDS

breast cancer; mathematical modelling; exponential model; primary tumor; secondary metastases; lymph nodes metastases; survival; predictor

ACM Reference format:
Ella Y. Tyuryumina and Alexey A. Neznanov. 2017. On Consolidated Predictive Model of the Natural History of Breast Cancer: Primary Tumor and Secondary Metastases in Patients with Lymph Nodes Metastases. In *Proceedings of DH '17, London, United Kingdom, July 2–5, 2017,* 7 pages.
https://doi.org/10.1145/3079452.3079461

DH '17, July 2–5, 2017, London, United Kingdom

ACM ISBN 978-1-4503-5249-9/17/07…$15.00.
https://doi.org/10.1145/3079452.3079461

1 INTRODUCTION

Breast Cancer (BC) is the most common cancer and also the leading cause of cancer mortality in women worldwide. BC accounts for about 20-25% of all cancer types in women [21].

Finding algorithms to predict the growth of tumors has piqued the interest of researchers ever since the early days of cancer research. Many studies were carried out as an attempt to obtain reliable data on the natural history of BC growth.

Mathematical modeling can play a very important role in the prognosis of BC. Various mathematical models were built to describe primary tumor (PT) growth and distant metastases (MTS) growth separately [19].

These days, an exponential, Gompertz, logistic and von Bertalanffy models are included in a group of classical mathematical models of PT growth [1]. For the breast data, the observed linear dynamics were best captured by an exponential model, which is situated for the description of PT growth and, also, for secondary distant MTS growth [7, 8, 10–16, 18]. As for Gompertz and logistic models, they are used rarely in order to describe PT growth or secondary distant MTS growth [5, 9, 17, 25].

The duration of the period from the first BC cell to death refers to the natural history of BC [2]. Secondary distant MTS appear in various time in different organs. The interval between removal of PT and the first clinical manifestation of MTS (MTS free survival time or non-visible period) determined by PT size, the number of affected lymph nodes and MTS growth rate [3, 5, 7–10, 12–16, 18, 22, 25, 26]. Survival (lifetime) is the period between the date of diagnosis (TNM staging system of BC) and the date of a patient death [21]. Survival among BC patients (%) indicates the percentage of people in a study or treatment group who are alive for a given period of time after diagnosis. The percentage of patients who live at least 5-, 10-, 15-, 20-, 25- and 30-years after being treated PT is defined as 5-, 10-, 15-, 20-, 25- and 30-years observed survival rate of BC patients [9, 21].

It is important to highlight that the natural history of BC continues after removal of PT. The next stage begins with secondary distant MTS manifestation. When the MTS reach the threshold volume, patients die from progression of BC [1, 4, 7–10, 12–16, 18, 20, 22]. All BC patients get a comprehensive treatment of PT, so the *whole natural history of BC* should include the period of secondary distant MTS growth (Fig. 1):

1) the **non-visible** period of PT growth;
2) the **visible** period of PT growth, diagnostics and removal of PT;
3) the **non-visible** period of secondary distant MTS growth;
4) the **visible** period of secondary distant MTS growth, diagnostics, treatment and patient's death.

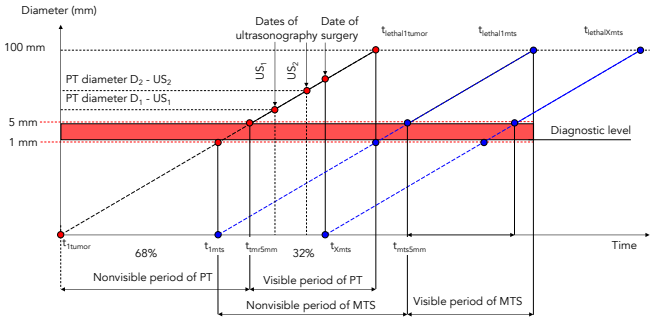

Figure 1: Scheme of the *whole natural history* of BC (stage I-II)[4, 7, 12, 14, 20, 22]. As it should be highlighted, the main feature is that model describes PT growth and secondary distant MTS growth as a whole (as indivisible dependent process). Ordinate (Y): Diameter of tumor (mm). Abscissa (X): Time (years).

Legend of Fig. 1:

t_{1tmr} — date of appearance of the first BC stem cell;

$t_{1tmr5mm}$ — date of appearance of the visible PT with size 5 mm;

$t_{lethal1tumor}$ — date of appearance of the lethal PT with size 100 mm (when PT reaches the threshold volume);

t_{1mts} — date of appearance of the first MTS stem cell, which coincides with the period of 20^{th} doubling time;

$t_{1mts5mm}$ — date of appearance the first visible MTS of breast cancer with size 5 mm;

$t_{lethal1mts}$ — date of appearance the first lethal MTS of BC with size 100 mm (when secondary distant MTS reaches the threshold volume);

t_{Xmts} — date of appearance nXmts cell of BC MTS, which coincides with date of surgery;

$t_{Xmts5mm}$ — date of appearance nXmts visible BC MTS with size 5 mm;

$t_{lethalXmts}$ — date of appearance nXmts lethal BC MTS with size 100 mm;

US_1 — date and sizes of the first US of PT;

US_2 — date and sizes of the second US of PT.

Given the relation between PT and MTS, the problem of discovering BC process seems to be twofold: firstly, it is important to describe the *whole natural history of BC* to understand the process as a whole; secondly, it is necessary to predict the period of a clinical MTS manifestation. Yet, the papers available for this do not offer mathematical growth models of MTS that relate to TNM classification. That leads to the demand building a mathematical model that rests on an exponential classical mathematical model and describes *whole natural history of BC* and corresponds to TNM classification. Moreover, the latter aspect of the problem is reflected only by statistical tools that are available as open source. In other words, a patient provides diagnostic data to predictor, and the tool calculates MTS free period and survival according to statistical data. Consequently, it is necessary to create a predictor that makes prognosis of BC for a patient independently from statistical data, and requires no expensive diagnostic data. Thus, this research possesses a novelty since it is the first time the following tools for BC have been proposed: a) *whole natural history of BC*; b) mathematical growth model corresponding to TNM; c) non statistical software tool for prediction of BC developing.

To avoid terminological ambiguities, we dwell upon recalling some standard terms and TNM staging system of BC (Table 1) [21].

Table 1: TNM staging system

Stage	Parameter T	Parameter N	Parameter M
I	T1	N0	M0
II	T1, T2	N0, N1	M0
III	T1, T2, T3, T4	N1, N2, N3	M0
IV	any T	any N	M1

Legend of Table 1:

parameter T — size of PT: T1 = 0.1 d ≤ 2 cm; T2 = 2 d ≤ 5 cm; T3 = d > 5 cm; T4 = spread;

parameter N — the number of affected lymph nodes: N0: n = 0; N1: n = 1-3; N2: n = 4-9; N3: n = 10-18;

parameter M — existence of distant MTS (lungs, bones, liver, etc): M0 = MTS not exist; M1 = MTS exist.

The **goal** of the research is to improve the prediction accuracy of BC process using the original **Co***nsolidated mathematical growth* **M***odel of primary tumor and secondary MTS of patients with lymph nodes MTS* (CoM-III). To make precise the scope of the study it is necessary to fulfil several **tasks**:

(1) modelling the *whole natural history* of PT and MTS for stage III;
(2) developing the adequate and precise CoM-III that reflects relations between PT and secondary MTS of patients with lymph nodes MTS;
(3) analysing the CoM-III scope of applications;
(4) implementing the model as a software tool.

Practical value. As it turns out, a new software tool for prediction of BC developing can calculate more accurately: a) MTS free period; b) survival for stage III of BC including PT and secondary MTS of patients with lymph nodes MTS. Moreover, the predictor can estimate a quality of treatment which was prescribed to a patient. Summarising: the CoM-III describes correctly PT and secondary distant MTS growth of T1N1M0, T2N1M0, T1-2N2M0, T3N1-2M0, T4N1-2M0, T1-4N3M0 [T1-4N1-3M0] stages in patients with lymph nodes MTS (N1-3).

2 MATERIALS AND METHODS

Consolidated mathematical growth model of PT and secondary MTS, I-II stages (CoMPaS)

In 2015 we proposed a consolidated mathematical growth model of PT and secondary MTS (CoMPaS) that describes correctly PT growth (parameter T) as well as secondary MTS growth (parameter M), corresponds to TMN [23]. Also, the CoMPaS might facilitate the survival (lifetime) and, as a consequence, make predictions of a future metastatic manifestation after removal of PT.

It is important to define several admissions, where rows 1-3 rests on [1, 7, 8, 10–15, 18, 20, 22, 24]:

1) an exponential growth model is used widely for description *natural* growth rate of the primary BC;
2) the *natural* rate of secondary distant MTS is the same as *natural* growth rate of the primary BC;
3) the period of appearance of the first metastatic cell of secondary distant MTS coincides with the 20th doubling of the primary BC. It allows us to define the **non-visible** growth period of MTS and the initial period of **non-visible** MTS manifestation;
4) the *whole nature history* of the PT and secondary distant MTS is 60 doublings (Fig. 2).

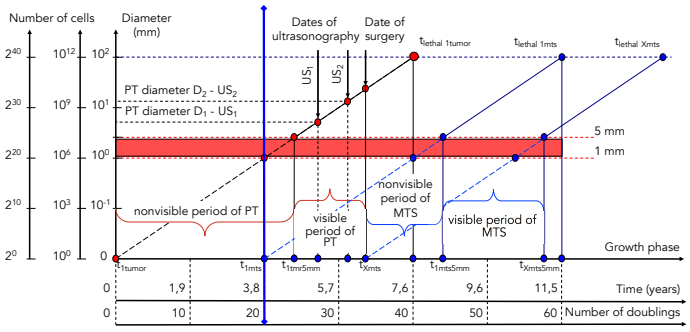

Figure 2: The first MTS cell appears on the 20th doubling of PT

Legend of Fig. 2:

red points	correspond to PT growth;
blue points	correspond to distant MTS growth;
the non-visible period	of MTS growth starts with a removal of PT;
red horizontal block	indicates a minimal size of tumor that can be diagnosed [1mm; 5mm];
growth rate	is equal as for PT as for MTS (CoMPaS for Stages I-II);
three vertical lines	on the left side of the Fig. 2 show the number of cells in tumor with corresponding diameter;
60th doubling	means the death of a patient;
$\xrightarrow{time}$	illustrates a mean survival at corresponded doubling.

The CoMPaS rests on an exponential growth model and consisted of nonlinear and linear determined equations [1, 3, 5, 7, 8, 10–16, 18, 20, 22, 24, 26]:

$$\begin{cases} \dfrac{dV}{dt} = \dfrac{\log 2}{DT} V, \quad t \le DT \ \log_2\left(\dfrac{\theta\ DT}{log2}\ V_0\right); \\ \dfrac{dV}{dt} = \theta \log\ V, \quad t > DT \ \log_2\left(\dfrac{\theta\ DT}{log2}\ V_0\right); \\ V(t=0) = V_0 \\ Survival = PT_{\log(V)} + Nonvis_{log} + Vis_{log} = 60; \\ TVDT_{non} = TVDT_{vis} = \dfrac{NonVis_{days} + Vis_{days}}{NonVis_{log} + Vis_{log}}; \end{cases}$$

$\frac{\log 2}{DT}$	—	the fraction of proliferative cells times;
θ	—	drives the linear phase;
$pT_{\log(V)}$	—	the number of PT doublings;
$Nonvis_{log}$	—	the number of doublings for **non-visible** growth period of MTS;
Vis_{log}	—	the number of doublings for **visible** growth period of MTS;
$TVDT$	—	tumor volume doubling time;
60 doublings	—	the *whole nature growth history* of the PT and secondary distant MTS.

According to M. Schwartz (1961): "the doubling time (DT), representing the time for 1 cell (or of the tumor as a whole if each cell has the same doubling time) to double in volume, and it is equivalent to the interval between successive mitoses" [20]:

$$DT = \frac{\log 2\ (t_1 - t_0)}{\log V_1 - \log V_0},$$

DT	—	the period of doubling time;
V_1	—	tumor volume at time t of the pre-surgery measurement;
V_0	—	tumor volume at time of the first measurement;
$t_1 - t_0$	—	the period between the first and pre-surgery measurements (days).

3 RESULTS

Consolidated mathematical growth model of PT and secondary MTS of patients with lymph nodes MTS, III stage (CoM-III)

Stage III (T1-4N1-3M0) means that lymph nodes MTS (N1-3) exist meanwhile PT (T1-4) is growing [6, 21]. Moreover, patients with lymph nodes MTS have lower survival comparing with patients without lymph nodes MTS. Unfortunately, the papers available for this do not offer mathematical growth models of stage III of BC that cover growth process of PT and secondary MTS in patients with lymph nodes MTS.

We propose a new mathematical growth model for PT and secondary MTS in patients with lymph nodes MTS. The model may help to improve predicting accuracy of BC process using an original mathematical model referred to CoM-III and corresponding software. Consequently, we are interested in:

1) modelling the *whole natural history* of PT and secondary MTS in patients with lymph nodes MTS;
2) developing adequate and precise CoM-III that reflects relations between PT and MTS;
3) analysing the CoM-III scope of application; 4) implementing the model as a software tool.

The period of appearance of the first metastatic cell coincides with 20^{th} doubling of PT. At stage III the period of MTS manifestation depends on the number of lymph nodes MTS.

The CoM-III rests upon CoMPaS, and by complementing formulas::

$$\begin{cases} sMts_{\log(V)(N+)} = pT_{\log(V)} - 20; \\ K_{sMts(N+)} = \dfrac{20}{sMts_{\log(V)(N+)} \times 18 \times n} + 1; \\ TVDT_{sMts(N+)} = \dfrac{TVDT_{pT}}{K_{sMts(N+)}}; \end{cases}$$

$sMts_{\log(V)(N+)}$ — the number of doublings of secondary MTS in patients with lymph nodes MTS;

$pT_{\log(V)}$ — the number of doublings of PT;

20 — the number of doublings of PT that coincides with the appearance of the first metastatic cell;

$K_{sMts(N+)}$ — correcting coefficient of secondary MTS of patients with lymph nodes MTS relating with PT growth rate;

n — the number of affected lymph nodes (min = 0, max = 18);

$TVDT_{sMts(N+)}$ — tumor volume doubling time of secondary MTS of patients with lymph nodes MTS relating with tumor volume doubling time of the PT ($TVDT_{pT}$).

It allows us to calculate different growth periods of PT and secondary MTS of patients with lymph nodes MTS:

1) **non-visible** period for PT;
2) **non-visible** period for secondary MTS of patients with lymph nodes MTS;
3) **visible** period for secondary MTS of patients with lymph nodes MTS.

Predictor CoM-III

At this stage, it is relevant to shed light on predictor specifications. We implement the CoM-III as a software tool. The application is build using Swift and referred as CoMPaS. The CoMPaS is available for iOS devices (iOS 9+).

INPUT DATA:

- the first ultrasound diagnostic data:
  - date
  - diameter (mm)
- the second ultrasound diagnostic data:
  - date
  - diameter (mm)
- the number of affected lymph nodes (n)

OUTPUT DATA:

- forecast:
  - the number of months
  - category of forecast:
    - favorable
    - mid-favorable
    - unfavorable

To flesh this out, the Fig.3 provides a clinical example for a patient with seven affected lymph nodes.

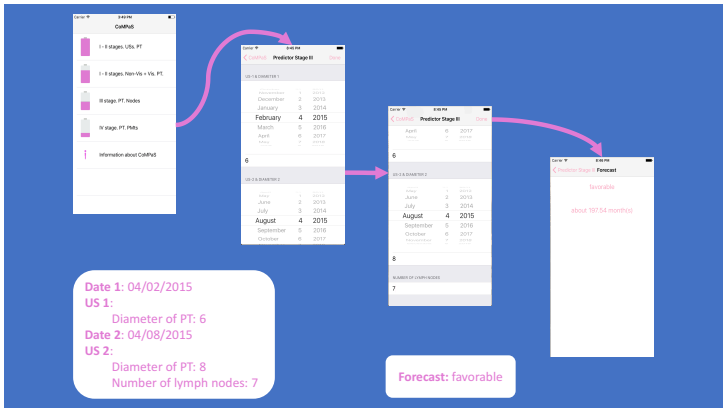

Figure 3: Clinical example

Under such circumstances, it is necessary to collect predictions in one database to compare forecasts with real data and estimate effectiveness of proposed model. Consequently, the CoMPaS connects to database that allows us to test application and model.

As it turns out, the new predictive tool: 1) is a solid foundation to develop future studies of BC models; 2) does not require any expensive diagnostic tests; 3) is the first predictor that makes forecast using only current patient data, whilst the others are based on the additional statistical data.

Calculations of *whole natural history* resting on CoM-III

Given all above, it is relevant to dwell upon building the *whole natural history* of BC stage III. Provided formulas allow calculating:

- the number of doublings for secondary MTS relying only on two measurements of PT sizes;
- the correcting coefficient of secondary MTS spreading rate in patients with lymph nodes MTS relating with PT growth rate;
- doubling time of secondary MTS.

Data of a mean diameter of PT for each stage (T1, T2, T3, T4) is obtained from table 1 of paper J. Engel et al. [6]. The number of affected lymph nodes corresponds to TNM staging system of BC [21]. The variety of doubling time of PT (T1, T2, T3, T4) relying on PT sizes is calculated from fig 6 of paper D. Holzel et al. [9]. Table 2 shows results of calculations via CoM-III.

Table 2: T1-3N0-3M0

		T1a (mm) $1 < d \leq 5$	T1b (mm) $5 < d \leq 10$	T1c (mm) $10 < d \leq 20$	T2 (mm) $20 < d \leq 50$	T3-4 (mm) $d > 50$	
1	Mean size of PT at surgery	4.5	8.5	15.1	28.5	64.6	[6]
2	$pT_{\log}$	26.4	29.2	31.7	34.4	38.0	
3	$TVDT_{pT}$	80.0	75.0	70.0	65.0	60.0	[9]
4	Mean $K_{sMts(N-)}$ (N0, n=0)	1.00	1.00	1.00	1.00	1.00	
5	$TVDT_{sMts(N-)}$ (N0, n=0)	80.0	75.0	70.0	65.0	60.0	[23]
6	Mean $K_{sMts(N+)}$ (N1, n_{mean}=2)	1.23	1.16	1.12	1.09	1.05	
7	$TVDT_{sMts(N+)}$ (N1, n_{mean}=2)	65.09	64.44	62.37	59.77	56.90	[23]
8	Mean $K_{sMts(N+)}$ (N2, n_{mean}=6)	1.69	1.49	1.37	1.26	1.16	
9	$TVDT_{sMts(N+)}$ (N2, n_{mean}=6)	47.42	50.28	51.21	51.48	51.56	[23]
10	Mean $K_{sMts(N+)}$ (N3, n_{mean}=14)	2.60	2.15	1.86	1.61	1.38	
11	$TVDT_{sMts(N+)}$ (N3, n_{mean}=14)	30.74	34.93	37.71	40.30	43.42	[23]

Legend of Table 2:

- TNM parameters depend on PT size: T1, T2, T3, T4, N1, N2, N3, M0 [21] (see detailed description on Fig. 5 - Fig. 9);
- row 1 uses data of the mean sizes of PT at surgery from tables of paper [6];
- row 2 is calculated from row 1;
- row 3 uses data from figure of paper [9];
- n_{mean} is the mean number of lymph nodes whereas:
  - N1 means that the number of lymph nodes can equal any integer number from compact [1; 3];
  - N2 means that the number of lymph nodes can equal any integer number from compact [4; 9];
  - N3 means that the number of lymph nodes can equal any integer number from compact [10; 18];
  - N1-3 means that the number of lymph nodes can equal any integer number from compact [1; 18].
- $K_{sMts(N-)}$ means no MTS in lymph nodes (N0, n_{mean}=0);
- the variety of $K_{sMts(N+)}$ is calculated for N1-3;
- the variety of $TVDT_{sMts(N+)}$ is calculated for N1-3.

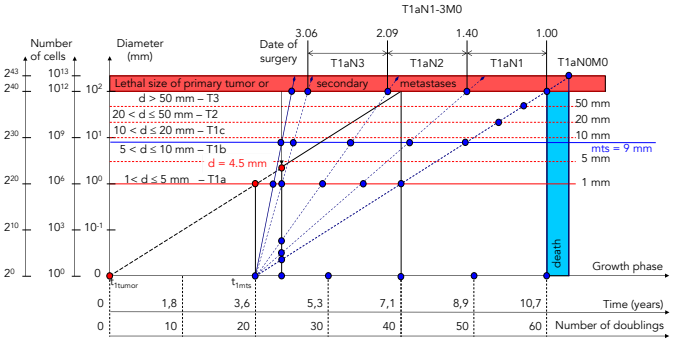

Figure 4: T1aN0-3M0. *Whole natural history* of PT and secondary MTS of patients with lymph nodes MTS according to CoM-III. Parameter T (T1a: 1mm < d ≤ 5mm) - diameter of PT. Parameter N (N0: n=0; N1: n=1-3; N2: n=4-9, N3: n=10-18) - affected lymph nodes.

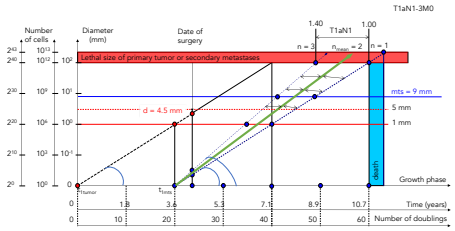

Figure 5

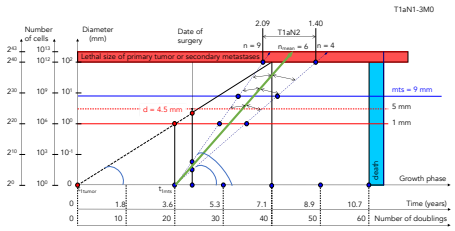

Figure 6

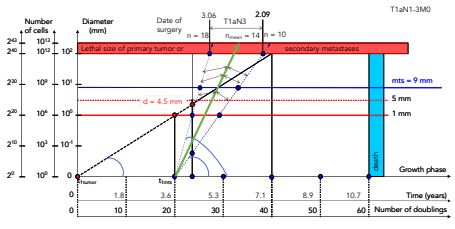

Figure 7

Legend of Fig. 5-7:

red points correspond to PT growth;
blue points correspond to distant MTS growth;
- - lines define boundaries of correcting coefficients values for patients with N1, N2, N3;
green line corresponds to the mean number of lymph nodes n_{mean}.

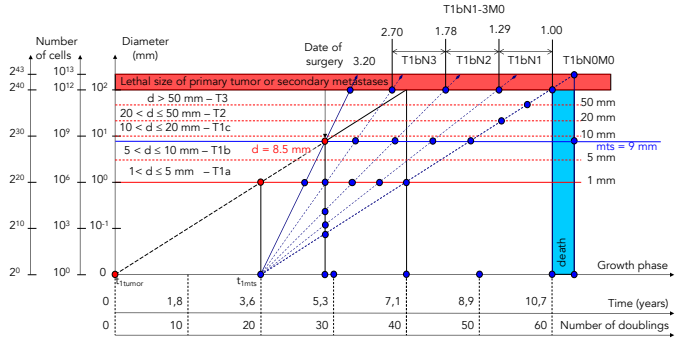

Figure 8: T1bN0-3M0. *Whole natural history* of PT and secondary MTS of patients with lymph nodes MTS according to CoM-III. Parameter T (T1b: 5mm < d ≤ 10mm) - diameter of PT. Parameter N (N0: n=0; N1: n=1-3; N2: n=4-9, N3: n=10-18) - affected lymph nodes.

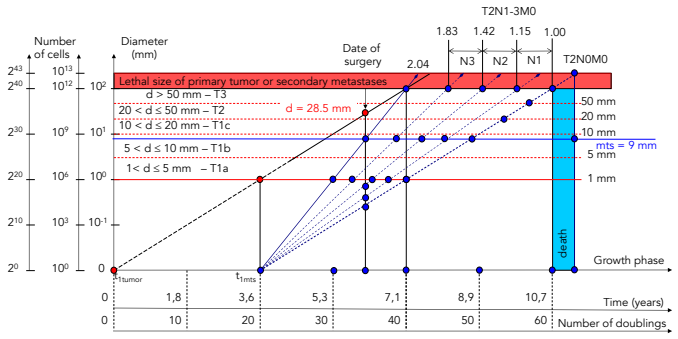

Figure 10: T2N0-3M0. *Whole natural history* of PT and secondary MTS of patients with lymph nodes MTS according to CoM-III. Parameter T (T2: 20mm < d ≤ 50mm) - diameter of PT. Parameter N (N0: n=0; N1: n=1-3; N2: n=4-9, N3: n=10-18) - affected lymph nodes.

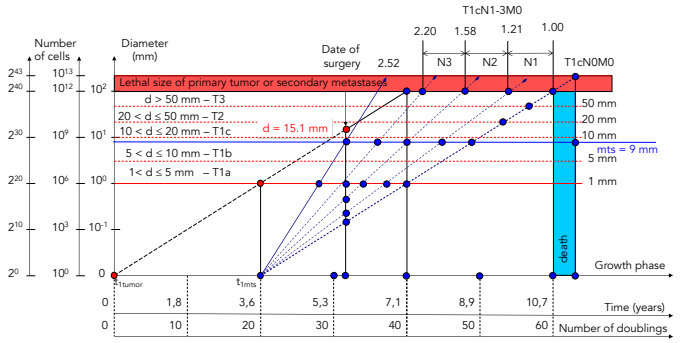

Figure 9: T1cN0-3M0. *Whole natural history* of PT and secondary MTS of patients with lymph nodes MTS according to CoM-III. Parameter T (T1c: 10mm < d ≤ 20mm) - diameter of PT. Parameter N (N0: n=0; N1: n=1-3; N2: n=4-9, N3: n=10-18) - affected lymph nodes.

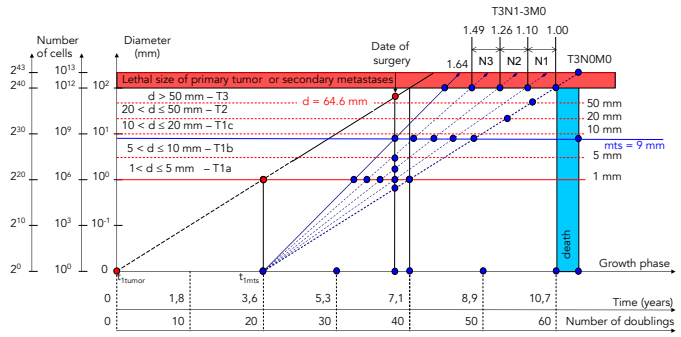

Figure 11: T3N0-3M0. *Whole natural history* of PT and secondary MTS of patients with lymph nodes MTS according to CoM-III. Parameter T (T3: d > 50mm) - diameter of PT. Parameter N (N0: n=0; N1: n=1-3; N2: n=4-9, N3: n=10-18) - affected lymph nodes.

4 CONCLUSION

The CoM-III model and predictive software: a) detect different growth periods of PT and secondary distant MTS growth in patients with lymph nodes MTS; b) make forecast of the period of secondary distant MTS appearance in patients with lymph nodes MTS; c) have higher average prediction accuracy than the other tools; d) can improve forecasts on survival of BC and facilitate optimisation of diagnostic tests. The following are calculated by CoM-III: the number of doublings for **non-visible** and **visible** growth periods of secondary distant MTS; tumor volume doubling time (days) for **non-visible** and **visible** growth periods of secondary distant MTS. The original CoM-III enables us, for the first time, to predict the *whole natural history* of PT and secondary distant MTS growth on each stage (T1, T2, T3, T4) for patients with/without lymph nodes MTS relying only on PT sizes. Summarising: CoM-III a) describes correctly PT and secondary distant MTS growth of T1N1M0, T2N1M0, T1-2N2M0, T3N1-2M0, T4N1-2M0, T1-4N3M0 [T1-4N1-3M0] stages in patients with lymph nodes MTS (N1-3); b) facilitates the understanding of the appearance period and inception of secondary distant MTS.

Work still to be done: 1. To test the CoM-III on clinical data. 2. To analyse forecasts statistically. 3. To implement CoM-III to medical practice.

REFERENCES

[1] S. Benzekry, C. Lamont, A. Beheshti, J.M. Tracz, A.and Ebos, L. Hlatky, and P. Hahnfeldt. 2014. Classical mathematical models for description and prediction of experimental tumor growth. *PLoS Comput Biol* 10, 8 (aug 2014). DOI: https://doi.org/10.1371/journal.pcbi.1003800

[2] H.J. Bloom, W.W. Richarson, and E.J. Harries. 1962. Natural history of untreated breast cancer (1805-1933). Comparison of untreated and treated cases according to histological grade of malignancy. *Br Med J* 2, 5299 (1962), 213–221. https://www.ncbi.nlm.nih.gov/pubmed/13870135

[3] S.E. Clare, F. Nakhlis, and J.C. Panetta. 2000. Molecular biology of breast cancer metastasis. The use of mathematical models to determine relapse and to predict response to chemotherapy in breast cancer. *Breast Cancer Res* 2, 6 (2000), 430–435. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC138666/

[4] V.P. Collins, R.K. Loeffler, and H. Tivey. 1956. Observations on growth rates of human tumors. *Am J Roentgen* 76, 5 (1956), 988–1000. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2008056/

[5] F.A. Coumans, S. Siesling, and L.W. Terstappen. 2013. Detection of cancer before distant metastasis. *BMC Cancer* 13, 283 (jun 2013), 1–12. DOI: https://doi.org/10.1186/1471-2407-13-283

[6] J. Engel, R.T. Emeny, and D. Holzel. 2012. Positive lymph nodes do not metastasise. *Cancer Metastasis Rev* 31, 1-2 (2012), 235–246. DOI: https://doi.org/10.1007/s10555-011-9343-7

[7] S. Friberg. 2005. On the growth rates of human malignant tumors: implications for medical decision making. *J Oncol* 55, 1 (aug 2005), 1–22. DOI: https://doi.org/10.1002/(SICI)1096-9098(199708)65

[8] P.M. Gullino. 1977. Natural history of breast cancer. Progression from hyperplasia to neoplasia as predicted by angiogenesis. *Cancer* 39, 6 (jun 1977), 2697–2703. DOI: https://doi.org/10.1002/1097-0142(197706)39

[9] D. Holzel, R. Eckel, R.T. Emeny, and J. Engel. 2010. Distant metastases do not metastasise. *Cancer Metastasis Rev* 29, 4 (sep 2010), 737–750. DOI: https://doi.org/10.1007/s10555-010-9260-1

[10] S. Koscielny, M.G. Le, and M. Tubiana. 1989. The natural history of human breast cancer. The relationship between involvement of axillary lymph nodes and the initiation of distant metastases. *Br J Cancer* 59, 5 (may 1989), 775–782. https://www.ncbi.nlm.nih.gov/pubmed/2736212

[11] S. Koscielny, M. Tubiana, and M.G. Le. 1984. Breast cancer: relationship between the size of the primary tumour and the probability of metastatic dissemination. *Br J Cancer* 49, 6 (jun 1984), 709–715. https://www.ncbi.nlm.nih.gov/pubmed/6733019

[12] S. Koscielny, M. Tubiana, and A.-J. Valleron. 1985. A simulation model of the natural history of human breast cancer. *Br J Cancer* 52, 4 (oct 1985), 515–524. https://www.ncbi.nlm.nih.gov/pubmed/4063132

[13] Y.T. Lee and J.S.Jr. Spratt. 1972. Rate of growth of soft tissue metastases of breast cancer. *Cancer* 29, 2 (feb 1972), 344–348. DOI: https://doi.org/10.1002/1097-0142(197202)29

[14] R.S. Lin and S.K. Plevritis. 2012. Comparing the benefits of screening for breast cancer and lung cancer using a novel natural history model. *Cancer Causes Control* 23, 1 (jan 2012), 175–185. DOI: https://doi.org/10.1007/s10552-011-9866-9

[15] V.M. Moiseenko. 2002. Natural history of breast cancer. *Practical oncology* 1 (2002), 6–14. http://www.practical-oncology.ru/arh009/02.pdf

[16] R. Molina-Pena and M.M. Alvarez. 2012. A simple mathematical model based on the cancer stem cell hypothesis suggests kinetic commonalities in solid tumor growth. *PLoS One* 7, 2 (feb 2012), 175–185. DOI: https://doi.org/10.1126/science.1065467

[17] L. Norton. 1988. A Gompertzian model of human breast cancer growth. *Cancer Res* 48, 24 (1988), 7067–7071. DOI: https://doi.org/10.1126/science.1065467

[18] A.W. Pearlman. 1976. Breast cancer: Influence of growth rate on prognosis and treatment evaluation. *Cancer* 38, 4 (1976), 1826–1833. https://www.ncbi.nlm.nih.gov/pubmed/991096

[19] I.A. Rodriguez-Brenes, N.L. Komarova, and D. Wodarz. 2013. Tumor growth dynamics: insights into evolutionary processes. *Cancer* 28, 10 (oct 2013). DOI: https://doi.org/10.1126/science.1065467

[20] M. Schwartz. 1961. A biomathematical approach to clinical tumour growth. *Cancer* 14 (nov 1961), 1272–1294. DOI: https://doi.org/10.1002/1097-0142(196111/12)14

[21] R. Shah, K. Rosso, and S.D. Nathanson. 2014. Pathogenesis, prevention, diagnosis and treatment of breast cancer. *World J Clin Oncol* 5, 3 (aug 2014), 283–298. https://www.ncbi.nlm.nih.gov/pubmed/25114845

[22] M. Tubiana and A. Courdi. 1989. Cell proliferation kinetics in human solid tumors: relation to probability of metastatic dissemination and long-term survival. *Radiother Oncol* 15, 1 (1989), 1–18. https://www.ncbi.nlm.nih.gov/pubmed/2664909

[23] Ella Turumina and Alexey Neznanov. 2015. Consolidated mathematical growth model of breast cancer. In *School and Conference of the Kharkevich Institute for Information Transmission Problems Information Technologies and Systems*. IITP RAS, http://itas2015.iitp.ru/pdf/1570162553.pdf, 38–51. arXiv:http://itas2015.iitp.ru/pdf/1570162553.pdf

[24] L. von Bertalanffy. 1957. Quantitative laws in metabolism and growth. *Q Rev Biol* 32, 3 (1957), 217–231. https://www.ncbi.nlm.nih.gov/pubmed/13485376

[25] H. Weedon-Fekjaer, B.H. Lindqvist, L.J. Vatten, O.O. Aalen, and S. Tretli. 2008. Breast cancer tumor growth estimated through mammography screening data. *Breast Cancer Rev* 10, 3 (2008), 1–13. DOI: https://doi.org/10.1186/bcr2092

[26] H.R. Withers and S.P. Lee. 2006. Modeling growth kinetics and statistical distribution of oligometastases. *Semin Radiat Oncol* 16, 2 (2006), 111–119. DOI: https://doi.org/10.1016/j.semradonc.2005.12.006