Use of electronic patient data storage for evaluating and setting the risk category of late effects in childhood cancer survivors

[1]Samuli Rajala, [1,2]Liisa S Järvelä, [1,2]Anu Huurre, [1,2]Marika Grönroos, [3]Päivi Rautava, [1,2,4]Päivi M Lähteenmäki

[1]University of Turku, Turku, Finland

[2]Department of Pediatrics and Adolescent Medicine, Turku University Hospital, Turku, Finland

[3] University of Turku, Public Health and Turku University Hospital, Research Services, Turku, Finland

[4] Swedish Childhood Cancer Registry, Karolinska Institutet, Stockholm, Sverige

Corresponding author:

Liisa S. Järvelä, MD, PhD

Pediatric Hematology/Oncology Division

Department of Pediatrics, Turku University Hospital,

Kiinamyllynkatu 4-8, FI-20520 Turku, Finland

E-mail: liisa.jarvela@utu.fi

Word counts: a) Abstract 249 b) Main text 2883

Number of Tables: 4

Number of Figures: 1

Running title: Automatized late effect risk categorization

Keywords: algorithm, childhood cancer survivor, electronic tool, late effect, long-term follow-up

CCS	Childhood cancer survivor
HSCT	Hematopoietic stem cell transplantation
CED	cyclophosphamide equivalent doses

ICD	International classification of diseases
CRP	C-reactive protein

ABSTRACT

Background: Many of the late effects of cancer treatment in childhood may occur even decades after the treatment, and only a minority of the survivors remain as healthy as their peers. Providing appropriate long-term care for childhood cancer survivors after transition to primary health care is a challenge. Both survivors and primary care providers need information on potential late effects. The lack of a systematic late effect follow-up plan may lead to excessive use of healthcare services or delayed intervention. While manual compilation of individual follow-up plans is time-consuming for experienced clinicians, electronic algorithms may be feasible.

Procedure: In Finland, international guidelines for determining the risk of late effects have been implemented. Nationally, Turku University Hospital was asked with developing an automatized system for calculating the risk of late effects, based on electronic patient records saved in the hospital data lake. An electronic algorithm that uses details from exposure-based health screening guidelines published by the Children's Oncology Group was created. The results were compared with those manually extracted by an experienced clinician.

Results: Significant concordance between the manual and algorithm-based risk classification was found. A total of 355 patients received a classification using the algorithm, and 325 of those matched with the manual categorization, producing a Cohen's coefficient of 0.91 (95% confidence interval 0.88–0.95).

Conclusion: Automated algorithms can be used to categorize childhood cancer survivors efficiently and reliably into late effect risk groups. This further enables automatized compilation of appropriate individual late effect follow-up plan for all survivors.

INTRODUCTION

Childhood cancer survivors (CCSs) with late effects comprise a significant new patient group in the healthcare system, and the number of childhood cancer survivors is constantly increasing. The current overall survival rate for childhood malignancies is approximately 80 %[1,2]. However, studies show that virtually all survivors develop at least one chronic health condition by the age of 45 years[3-5]. In many countries, individualized follow-up plans for CCSs are still not implemented. However, many patients could benefit from anticipatory guidance regarding health promotion and disease prevention aimed at minimizing the risk of future morbidity and mortality[6,7]. Cancer and its treatment during childhood or adolescence may have numerous different physical and psychosocial effects that may lead to excess morbidity or early mortality when compared to those in the general population. Essentially, any organ can be affected by chemotherapy, radiation, or surgery used for effective treatment.

The transition of CCSs from pediatric to adult healthcare poses a major challenge. Most late effects occur decades later, and recognition of the symptoms is often delayed in general practice[8-11]. To facilitate comprehensive and systematic follow-up of CCSs, the Children's Oncology Group has organized exposure-based health screening guidelines (http://www.survivorshipguidelines.org/). The length of treatment, high cumulative doses of chemotherapy and irradiation, multimodal therapy, and relapse therapy are associated with an increased risk of late effects.

Risk-based care, including a systematic plan for lifelong screening, follow-up, and prevention, is recommended for all cancer survivors. This plan should incorporate risks based on the cancer type, cancer therapy, possible genetic predispositions, lifestyle behaviors, and comorbid health conditions[6]. The classification of risk-based follow-up is determined by multiple factors, such as the patient's age at the time of cancer diagnosis, cancer histology, organs/tissues affected by cancer, as well as treatment modalities including surgical procedures, chemotherapeutic agents and their cumulative doses, irradiation doses and treatment fields, and possible hematopoietic stem cell transplantation (HSCT).

In a common effort between the national representatives of pediatric and adult oncology in Finland, the pre-existing international recommendations for determining the risk of late effects were adapted to serve as nation-wide guidelines for health-care authorities when planning late effect follow-ups for former patiens with childhood and young adult cancer[8,12]. After the statement of the National Working Group was published in 2014[10], each of the five Finnish University Hospital districts established a late effect clinic for their respective catchment area. In addition to the work performed at these five follow-up clinics, the role of a national developer was allocated to the Western Cancer Center of Finland hosted by Turku University Hospital within the Hospital District of Southwest Finland. One of the development projects involved creating an algorithm to calculate the late effect risk for each survivor based on data from the hospital-based electronic informatics center that gathers clinical data and outcome information in real-life settings (i.e., the hospital data lake). A description of the basis for tailoring the late effect follow-up plans is presented in Table 1. This shared-care model has been adapted from the original British initiative[12-14].

It has been estimated that it takes several hours for an experienced clinician to create a complete treatment summary from patient records[15]. In this study, we aimed to develop an electronic tool to calculate individual late effect risks for childhood cancer survivors based on their specific diagnosis and treatment details (surgery, doses of chemotherapy, doses and fields of irradiation, and possible HSCT). The goal of this initiative was to improve the quality of the follow-up of the CCSs and lighten the burden of experienced clinicians, who currently manually extract risk assignments for each patient individually.

SUBJECTS AND METHODS

The information stored in electronic patient files is a valuable data source, although it is often difficult to utilize in research. Auria Clinical Informatics carefully organizes, harmonizes and maintains the data in the data lake of the Hospital District of Southwest Finland and provides both research services and a secure data analysis platform for data-driven real-world analytical studies. The medical record data at our hospital district have been stored electronically since 2004; the data include details on demographics, inpatient periods and outpatient visits, clinical diagnoses and procedures, pathological diagnoses and reports, imaging results, chemotherapy and irradiation treatments, inpatient medications and outpatient prescriptions, laboratory measurements, and clinical narratives (Fig. 1). Currently, the data lake contains clinical data for approximately 1.5 million patients who have visited Turku University Hospital. The data are longitudinal, making it possible to follow individual patient trajectories and study outcomes. The data are pseudonymized, protecting the identity of the patients, while making it possible to link data elements to individual patients. Much of the electronic data are structured, and text mining can be used on unstructured clinical narratives when needed.

The target group for this study included patients diagnosed with childhood cancer (age 0–16 years) at the Pediatric and Adolescent Hematology/Oncology unit of Turku University Hospital after January 1, 2004, and whose treatment had ended by December 31, 2017. The criteria for recognizing patients from the hospital's data lake were International Classification of Diseases (ICD)-10 diagnosis codes in group C, and in group D from D32 to D43.

The details of the cancer treatments and possible additional diagnoses recorded since January 1, 2004 were categorized into three risk-defining groups, as presented in Table 2. Irradiation doses were categorized as follows: no radiotherapy, 0.1–19.9 Gy, 20 Gy or more. The cumulative doxorubicin equivalent anthracycline dose was categorized as follows: no anthracyclines, dose of less than 250 mg/m^2, dose of 250 mg/ m^2 or more. Regarding alkylating agents, a high dose was originally defined as dose of more than 3 000 mg/m^2 of cyclophosphamide or dose of more than 60 000 mg/m^2 of ifosfamide. We did not have a specific code for HSCT available for the entire study period, and thus surrogate parameters were used to identify survivors who had undergone HSCT (treatment with total-body irradiation or intravenous busulfan or melphalan). Neurological diagnosis of hemiplegia/ other paralysis, and a positive blood culture with a C-reactive protein (CRP) level greater than 200 mg/l were tested as surrogates for severe neurological problems or severe septicemias that could cause late sequelae in patient.

The process of clinical patient data analysis was as follows: The annual clinically produced patient lists with new cancer diagnoses from January 1, 2004 to December 31, 2017 at the Pediatric and Adolescent Hematology/Oncology unit in Turku University Hospital were merged, and all patients who had completed cancer therapy were included in the study

population. Patient records (including chemotherapy and irradiation charts) were then manually evaluated. The cumulative dose of each chemotherapy agent was calculated and recorded on the survivorship passport form (MG, AH, LJ, PL) in addition to all other treatment details or significant toxicities. An experienced clinician (PL) manually defined the late effect risk classification based on the details described in Table 1.

After the classification was performed using the algorithm, the sensitivity, specificity, and positive and negative predictive values were calculated for the algorithm. A quadratic weighted Cohen's kappa coefficient was calculated to test the agreement between the algorithm-based and manual risk group categorization. All statistical analyses were conducted using R 3.6.3 software (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

In total, 638 patients were identified from the data lake based on ICD-10 codes. Initially, 16 of them could not be classified due to non-existing health record information on any cancer treatment. The remaining 622 patients received algorithm-based classifications. The manually produced survivorship passports and classifications provided by the algorithm were then compared case by case.

The comparison between the algorithm-based classification and manually produced survivorship passports showed that the algorithm had detected 267 patients that were not on the clinician's list. After closer examination of these additional cases, the reasons for misclassification were determined as follows: patients whose treatment had ended before January 1, 2004 but had follow-up appointments registered as new cases (n=102); patients

who were treated elsewhere but had an appointment at Turku University Hospital (n=62); patients aged 16 years who were treated in the adult department (either for thyroid cancer or other adult-type malignancies; n=9); patients who were misdiagnosed with a tumor at the beginning of their diagnostic path (n= 26); additional patients who were falsely detected as patients with cancer, even though their C-code referred to a cardiovascular ATC-drug code registered as a cause of poisoning (n=13); patients who were still receiving active cancer care (n=15); and patients who had incomplete medical record information in the data lake, which led to incorrect classification (n=10). With these incorrectly classified patients removed from the equation, 355 patients were suitable for post-cancer treatment risk classification.

Among these 355 patients with cancer, we noticed that some patients were classified into an unnecessarily high-risk category because of the significantly low cyclophosphamide dose threshold or surrogates of severe septicemia (high CRP level with a positive blood culture). The algorithm was then corrected so that the cyclophosphamide high-risk dosage threshold was set to 6 000 mg/m^2, to ensure that the algorithm identifies patients who are the most at risk of infertility[16,17]. Cases of a positive blood culture with a high CRP level were removed from the algorithm. After these corrections, re-categorization with the algorithm was performed for the same 355 patients.

With the corrected algorithm, the final tabulation showed that 91.6 % (325/355) of the survivors were classified correctly. Nine (2.53 %) of the wrongly classified survivors had a higher risk value, and 21 cases (5.92 %) had a lower risk value compared to the clinician's assessment (Table 3). The sensitivity, specificity, and positive and negative predictive values of the algorithm for each risk group are shown in Table 4. Calculation of Cohen's coefficient of stability led to a Cohen's coefficient of 0.91 (95 % confidence interval 0.88–0.95).

The main reason for the nine risk values that were higher than the manual value was that the algorithm interpreted any three-month shortage of data lake information regarding administered intravenous chemotherapy during the primary treatment as relapse, which classified a patient as a high-risk candidate. In most cases of lower values, the reason was that the chemotherapy dose information critical to the classification was missing from the data lake for unknown reasons. In a few cases, the patients' irradiation therapy or HSCT was timed after the age of 16 years and thus (accidentally) excluded from the algorithm. For two patients with craniopharyngioma, stereotactic irradiation was not recognized as an additional high-risk criterion.

DISCUSSION

The process of creating an electronic algorithm for late effect risk categorization showed that using current tools/definitions, it is already possible to achieve more than 90% accurate results from retrospective data, compared with those achieved from manual categorization performed by an experienced clinician.

Previous studies have shown that as many as two-thirds of CCSs experience one or more chronic health conditions that can be severe or even life-threatening, and the prevalence of chronic health conditions is much higher than that among the general population[3-5]. This has led to the importance of transitioning from pediatric to adult care and continuous monitoring well beyond the pediatric age. Additionally, the survivors themselves often lack information about the treatment administered or possible complications they may experience because of their previous cancer experience.

Risk-based classifications with follow-up guidelines (survivorship passport) allow survivors to receive the best possible care from any physician. This would also help to maintain a balance between over-screening and underscreening. Over-screening may cause undue fear of unlikely late effects and lead to high medical costs by unnecessary screening of remotely plausible complications. In contrast, underscreening may lead to overlooking a potentially life-threatening complication and delaying the intervention, possibly causing permanent damage. Long-term follow-up guidelines incorporated into an algorithm allow for defining survivors who will need surveillance, and determining the most effective screening method, when and at what frequency the screening should be initiated, and the measures that should be implemented.

The level of long-term follow-up care for the patients has most often been planned based on the risk categories that have originally been defined in the publications by Wallace et al. (2001)[8] and Hudson et al. (2011)[12]. In the present study, these criteria were used, except for irradiation. As the risks of irradiation differ widely depending on the organs in the irradiation field as well as other treatments that may have been administered, we decided to decrease the level of irradiation dose compared to that given by the US colleagues[12]. As it seems that many survivors already need a high-risk level of follow-up care with 20 Gy of irradiation, this limit was chosen to enable the algorithm to detect these survivors for high-risk follow-up. However, even lower doses of irradiation to certain fields may be significant in terms of surveillance recommendations (e.g., breast cancer surveillance), and thus, any lower radiation doses were assigned to the intermediate-risk group.

In the present study, the high-risk cyclophosphamide and ifosfamide cut-offs were used instead of cyclophosphamide equivalent doses (CEDs). While the use of CEDs would have been the best way of defining the risk of e.g., infertility[18], we do not yet have a clear cut-off that helps us categorize based merely on the CED. In our current algorithm, we chose alkylator doses that were related to the accumulation of risk factors, enabling the algorithm to accurately detect survivors with the highest risk of morbidity leading to even annual follow-up recommendations. However, for clinical implementation, the algorithm will be further developed so that both the follow-up risk category and the cumulative doses of chemotherapy, irradiation doses, and radiotherapy fields are retrieved as a printout. At this point, the CEDs will also be included.

Digital surveillance programs have already been successfully launched[15], but due to the lack of comprehensive/suitable data lakes, there have been no previous attempts to create algorithms for automatized risk categorization. Our current algorithm utilizes the Turku University Hospital data lake, which stores and aligns all electronic health records generated at the hospital**.** This data lake enables automatized late effect risk assessments for all patients using widely acknowledged criteria. With this information, it is possible to track their health statuses individually, and follow-up can be performed based on individual treatment histories.

Previous survey-based studies have indicated that primary care providers are concerned about their own readiness to take responsibility for the follow-up of childhood cancer survivors[19]. Considering this uncertainty, automated and individualized follow-up plans would be an easy solution. Multiple models of care for CCSs have been implemented and studied[12,20,21]. The available data do not support a single follow-up model for all circumstances. In a perfect setting with unlimited resources, all survivors would be followed up by a survivorship

specialist throughout their lives. As that is not accomplished, survivors should be risk-stratified based on treatment exposures. Survivors with a low risk of late effects should receive a survivorship care plan from their oncologist, or in this case from an electronic algorithm. For those with a low risk of late effects, this plan can be implemented by the primary care provider, while those with a higher risk of late effects need closer examination and should continue to be followed up at a survivorship clinic[21].

Although late treatment effects can be anticipated in most cases based on therapeutic exposures, a patient's individual risk is modified by multiple factors. Currently, it may not be possible to include all individual factors in an algorithm. However, rapid development of technologies may lead to machine learning solutions that help us include many psychosocial risks and warning signs in an algorithm by text mining from the notes of nurses and doctors. Furthermore, nationally defined procedure codes are now being used to record HSCTs, further improving the concordance between the current algorithm and a clinician's discretion. In addition, as our study was based on a retrospective survey approaching the emergence of the data lake structure, many of the shortcomings within the data collection and integration have already been solved. Thus, a patient receiving a diagnosis today is likely to have very comprehensive data stored, enabling highly representative risk calculations to be performed by our algorithm.

The lack of electronic patient records and comprehensive data lakes remain a shortcoming in the applicability of such automatized algorithms in some countries. However, as the treatment of childhood cancers is mainly centralized to tertiary hospitals, data lakes established at these centers may yield sufficient information for late effect risk categorization.

With the current patient record applications merged in a data lake, information on treatment details, such as irradiation and doses of chemotherapy, are stored in a reliable and structured manner. This allows automatized algorithms to efficiently and reliably categorize CCSs into late effect risk groups. To use the full potential of electronic patient record solutions, it is essential to continue building hospital data lakes. Efforts should be made to implement automatized late effect algorithms to facilitate appropriate late effect follow-up plans for all CCSs without the extensive use of clinician resources.

ACKNOWLEDGEMENTS

Tommi Kauko (a data scientist at Auria Clinical Informatics, Turku, Finland) is greatly acknowledged for his work involving the electronic data at the hospital data lake.

CONFLICT OF INTEREST STATEMENT

None of the authors have any financial or other conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

REFERENCES:

318 1. Gatta G, Botta L, Rossi S, et al. Childhood cancer survival in europe 1999-2007: Results of
319 EUROCARE-5--a population-based study. Lancet Oncol 2014;15:35-47.

320 2. Madanat-Harjuoja LM, Pokhrel A, Kivivuori SM, Saarinen-Pihkala UM. Childhood cancer
321 survival in finland (1953–2010): A nation-wide population-based study. Int J Cancer
322 2014;135:2129-2134.

323 3. Oeffinger KC, Mertens AC, Sklar CA, et al. Chronic health conditions in adult survivors of
324 childhood cancer. N Engl J Med 2006;355:1572-1582.

325 4. Diller L, Chow EJ, Gurney JG, et al. Chronic disease in the childhood cancer survivor
326 study cohort: A review of published findings. J Clin Oncol 2009;27:2339.

327 5. Robison LL, Hudson MM. Survivors of childhood and adolescent cancer: Life-long risks
328 and responsibilities. Nat Rev Cancer. 2014;14:61.

329 6. American Academy of Pediatrics Section on Hematology/Oncology Children's Oncology
330 Group. Long-term follow-up care for pediatric cancer survivors. Pediatrics 2009;123:906-915.

331 7. Armstrong GT, Kawashima T, Leisenring W, et al. Aging and risk of severe, disabling,
332 life-threatening, and fatal events in the childhood cancer survivor study. J Clin Oncol
333 2014;32:1218.

334 8. Wallace WHB, Blacklay A, Eiser C, et al. Developing strategies for long term follow up of
335 survivors of childhood cancer. Br Med J 2001;323:271-274.

336 9. Lund LW, Schmiegelow K, Rechnitzer C, Johansen C. A systematic review of studies on
337 psychosocial late effects of childhood cancer: Structures of society and methodological
338 pitfalls may challenge the conclusions. Pediatr Blood Cancer 2011;56:532-543.

339 10. Taskinen M. Childhood and adolescent cancer was cured--how to support health in
340 adulthood? Duodecim 2014;130:2320-2330.

341 11. Mulder RL, Van Der Pal H, Levitt GA, et al. Transition guidelines: An important step in
342 the future care for childhood cancer survivors. A comprehensive definition as groundwork.
343 Eur J Cancer 2016;54:64-68.

344 12. Hudson MM, Landier W, Ganz PA. Impact of survivorship-based research on defining
345 clinical care guidelines. Cancer Epidemiol Biomarkers Prev 2011;20:2085-2092.

346 13. Eiser C. Follow-up after childhood cancer: Evaluation of a three-level model. Eur J
347 Cancer. 2006;42(18):3186-3190.

348 14. Edgar A. Can intensity of long-term follow-up for survivors of childhood and teenage
349 cancer be determined by therapy-based risk stratification? BMJ Open 2013;3:101136.

350 15. Haupt R, Essiaf S, Dellacasa C, et al. The 'Survivorship passport'for childhood cancer
351 survivors. Eur J Cancer 2018;102:69-81.

352 16. Jahnukainen K, Heikkinen R, Henriksson M, Cooper TG, Puukko-Viertomies L, Mäkitie
353 O. Semen quality and fertility in adult long-term survivors of childhood acute lymphoblastic
354 leukemia. Fertil Steril 2011;96:837-842.

355 17. Poganitsch-Korhonen M, Masliukaite I, Nurmio M, et al. Decreased spermatogonial
356 quantity in prepubertal boys with leukaemia treated with alkylating agents. Leukemia
357 2017;31:1460-1463.

18. Green DM, Liu W, Kutteh WH, et al. Cumulative alkylating agent exposure and semen parameters in adult survivors of childhood cancer: A report from the st jude lifetime cohort study. Lancet Oncol 2014;15:1215-1223.

19. Poplack DG, Fordis M, Landier W, Bhatia S, Hudson MM, Horowitz ME. Childhood cancer survivor care: Development of the passport for care. Nat Rev Clin Oncol 2014;11:740.

20. Heirs M, Suekarran S, Slack R, et al. A systematic review of models of care for the follow-up of childhood cancer survivors. Pediatr Blood Cancer 2013;60:351-356.

21. Song A, Fish JD. Caring for survivors of childhood cancer: It takes a village. Curr Opin Pediatr 2018;30:864-873.

LEGENDS

TABLE 1. Current basis for tailoring individual late effect follow-up plans for childhood cancer survivors in Finland. The model is based on treatment factors and has been adapted from international publications[12,13].

TABLE 2. Original definitions of late-effect risk categories for the algorithm.

TABLE 3. Tabulation of the risk categories set by the algorithm and by an experienced clinician.

TABLE 4. Sensitivity, specificity, and positive and negative predictive values of the algorithm for each late effect risk group.

381

382

383 FIGURE 1. Auria Clinical Informatics collects real-time clinical data for the Biobank, quality

384 registration purposes, and clinical research from several electronic patient record sources.

TABLE 1. Current basis for tailoring individual late effect follow-up plans for childhood cancer survivors in Finland. The Model is based on treatment factors and has been adapted from international publications[16,17].

RISK OF LATE-EFFECTS	FICAN-WEST MODEL OF FOLLOW-UP
LOW Treatment with surgery only or low risk chemotherapy (no alkylation agents, no bleomycin, no anthracyclines, no epipodophyllotoxins)	Survivorship passport with the follow-up plan created at the last visit to pediatric hematology/oncology clinic using BCB-Medical platform[a]. One visit to the nurse at the late-effect clinic after the age of 18 for a review of the plan. Thereafter the basic healthcare is responsible for the physical visits. Annual eHealt contacts with KAIKU® (questionnaires, messages)[b]
INTERMEDIATE Other than low or high risk treatment	Survivorship passport with the follow-up plan created at the last visit to pediatric hematology/oncology clinic using BCB-Medical platform[a]. First visit the nurse at the late-effect clinic after the age of 18 for a review of the plan. Thereafter, visits at least with five years intervals. The basic healthcare takes care of the planned examinations, but follow-up clinic helps with the specialist care consultations if needed. Annual eHealt contacts with KAIKU® (questionnaires, messages)[b]
HIGH Treatment with stem cell transplantation. Treatment with high-dose of anthracyclines or alkylating agents or irradiation dose 20Gy or more	Survivorship passport with the follow-up plan created at the last visit to pediatric hematology/oncology clinic using BCB-Medical platform[a]. First visit to the nurse at the late-effect clinic after the age of 18 for a review of the plan. Thereafter, even annual visits if needed in order to keep all planned specialist consultations going on, and to give psychosocial support. Specialist care will take place at local hospitals. Annual eHealt contacts with KAIKU® (questionnaires, messages)[b]
[a] https://www.bcbmedical.com/?lang=en [b] https://kaikuhealth.com/	

TABLE 2. Original definitions of late-effect risk categories for the algorithm.

Low risk	Medium risk	High risk
Only surgical treatment		
No irradiation	No high risk irradiation	Irradiation > 20 Gy/field or whole-body radiation
Low risk chemotherapy (no alkylating agents, no anthracyclines, no platinum compounds, no bleomycin)	Chemotherapy that is not in low or high risk category	Anthracyclines with doxorubicin-equivalent dose ≥ 250 mg/sqm. High dose alkylating agents (Cyclophosphamide[a] > 3 000 mg/sqm or Ifosfamide > 60 000 mg/sqm) Stem cell transplantation (code WW3, or therapy with Busulfan and/or Melphalan). Brain tumor with any chemotherapy
	Paralysis (IDC-10 G81-83).[b] Blood culture positive and CRP > 200 mg/l[b]	Relapse (surrogate criteria: chemotherapy brake over 3 months)

[a] Cyclophosphamide dose was corrected up to 6 000 mg/sqm as current literature shows that risk of male infertility increases after this threshold.
[b] This criterion was removed from the final algorithm

TABLE 3. Tabulation of the risk categories set by the algorithm and by an experienced clinician.

Risk gategory/ classification method	Manual low	Manual medium	Manual high
Algorithm low	129	6	5
Algorithm medium	2	74	10
Algorithm high	1	6	122

TABLE 4. Sensitivity, specificity, and positive and negative predictive values of the algorithm for each late effect risk group.

	Low Risk	Medium Risk	High Risk
Sensitivity	97.7 %	86.0 %	89.1 %
Specificity	98.6 %	95.5 %	96.8 %
Positive Predictive Value	92.1 %	86.0 %	94.6 %
Negative Predictive Value	98.6 %	95.5 %	93.3 %

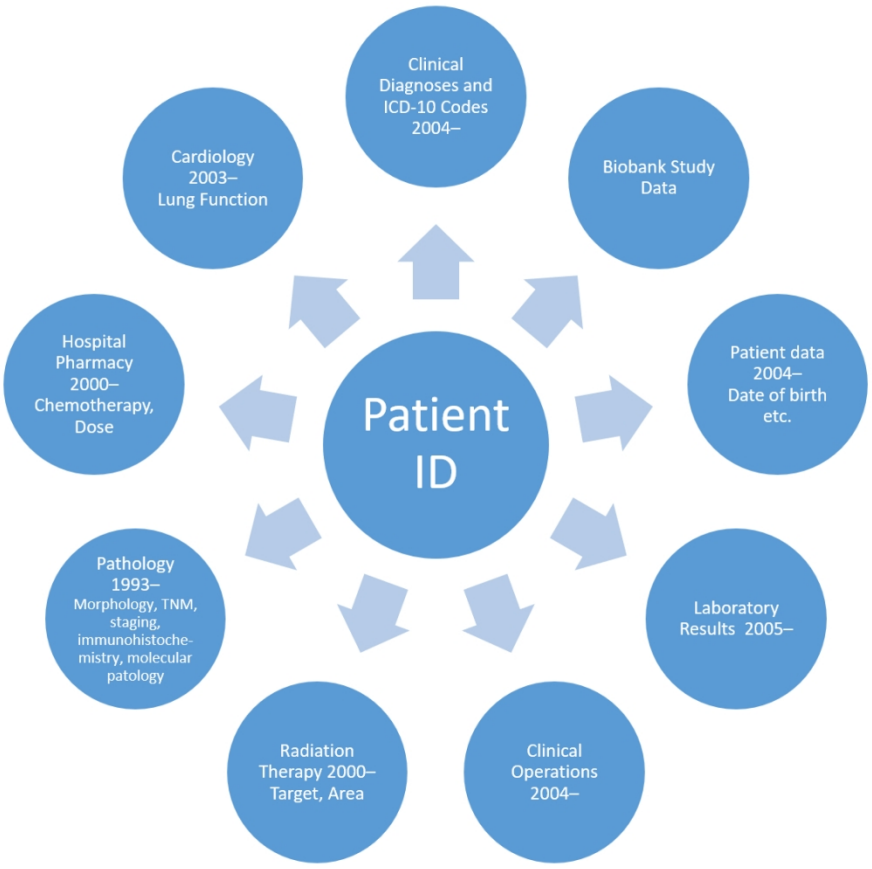

FIGURE 1. Auria Clinical Informatics collects real-time clinical data for the Biobank, quality registration purposes, and clinical research from several electronic patient record sources.

218x188mm (150 x 150 DPI)