

Evidence-based data and rare cancers: the need for a new methodological approach in research and investigation

Authors: S Mathoulin-Pélissier (1,2) and K Pritchard-Jones (3)

Affiliations

1 Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, Epicene team, UMR 1219, F-33000 Bordeaux, France

2 Clinical and Epidemiological Research Unit, INSERM CIC1401, Institut Bergonie, Comprehensive Cancer Center, F-33000 Bordeaux, France

3 University College London, UCL Great Ormond Street Institute of Child Health, 30 Guilford Street, London WC1N 1EH, U.K.

Abstract

Rare cancers are not so rare, their incidence is increasing and, as a group, they have worse survival than the common cancers. These factors emphasise the societal need to ensure sufficient focus on research into their biological basis, aetiological factors, new more effective therapies and organisation of healthcare to improve access to best practice and innovation. Accuracy of diagnosis is one of the first hurdles to be overcome, with around one third of tumours being reclassified – by type or risk group – when subject to a centralised pathology review process. Timely access to appropriate expert knowledge is a second challenge for patients – in Europe this is being addressed by the establishment of European Reference Networks (ERNs) as part of the EU cross border healthcare initiative. There are ERNs for adult solid and haematological cancers and childhood cancers, all of which are individually rare. These ERNs will facilitate creation of large databases of rare tumours that will incorporate knowledge of their molecular features and build an evidence base for the effectiveness of innovative, biology-directed therapies. With an increasing focus on ‘real world’ outcome data, research methodologies are evolving, to include randomised registry trials and data linkage approaches that exploit the ever-richer information held on patients in routine health care data. The inclusion of genomic analysis into cancer diagnosis, treatment and risk prediction raises many issues for the conduct of clinical research and cohort studies and personal data sharing. Sophisticated means of pseudonymisation, together with full involvement of affected and ‘at risk’ patients, are supporting novel research designs and access to data that will continue to build the evidence base to improve outcomes for patients with rare cancers.

Key words: rare cancer, evidence based-medicine, genomic, clinical trial, observational study, population based registry, international collaborations

1 Introduction

The increasing incidence and effects of cancer have led to greater use of cancer registries to understand the opportunities for intervention and the development of screening and research programs for cancer prevention. These initiatives have resulted in improvement of cancer detection, diagnostics, treatment, follow-up and research. However, these advances have not been applied to the same extent across all cancer types and patient groups. For rare cancers, because the number of patients is low, there is insufficient focus on accurate and timely diagnosis, effective treatment modalities and evidence based guidelines. Additionally, funds for research on rare cancers are limited and it is complex to perform clinical trials due to the lack of adequate sample sizes [1].

The principles of evidence based medicine (EBM) are relevant to all types of diseases including uncommon and rare cancers. In EBM, there is a rather clear hierarchy of evidence in which the highest form is the randomised controlled trial – the RCT. The randomisation assures that both known and unknown confounding factors are evened out between groups, and the control treatment assures that the intervention is different than the natural progression. Despite the strong evidence that they create, RCTs are very expensive and generally cumbersome. The logistics of an RCT are challenging, the recruitment process complicated, staff need to be trained and the study sites monitored. Observational studies on the other hand, are less expensive, but create weaker evidence. They are generally placed much further down the evidence ladder as, without randomisation, tests of efficacy are deemed less credible. However, they have the advantage that they recruit a much broader range of patients who more closely reflect the ‘real world’ of clinical practice. Through collection of detailed data on treatment and outcomes, such observational studies have the potential to demonstrate population impact of new interventions. Such an approach can be of particular interest in rare cancers.

Until recently, no universal definition of rare cancer existed. The RARECARE group from Europe proposed a practical definition of ‘rare’ as a cancer with an incidence rate of <6 cases per 100,00 population per year [2]. This group produced a list of clinically relevant, histologically defined cancers (almost 200). The rare cancer list proposed is based on the International Classification of Diseases for Oncology (ICD-O, 3rd version), the classification of tumours recognised worldwide. Rare tumour entities are relevant for clinical decision-making and clinical research, while families of tumours are relevant for organisation of health care. In the era of molecular targeted therapies, the molecular profile will also be relevant. Indeed, genetic and molecular profiling of common cancers can partition these into rarer subgroups and international agencies that preside over such classifications are constantly updating them. However, these ‘rare’ subgroups of commoner cancers often have mechanistic evidence for a therapy based on the molecular target that defines them. Hence,

their therapeutic and research needs are somewhat different to the individually rare cancers. The key figures of rare cancer burden in EU during 2000-07 were the following: 24% of all new malignancies and worse survival than common cancers [3]. In the US, approximately 20% of patients were diagnosed with rare cancers and 70% occurred in children and adolescents [4]. The worse 5-year relative survival (55% for rare cancers vs 75% for common during 2009-13 in USA and 49% vs 63.% during 2000-07 in EU) is more pronounced in adult patients and probably linked to a more advanced stage in rare cancers (59% 5-year relative survival for rare cancers vs 45% for common in USA)[3,4].

These data provide a clear picture of the global situation for rare cancers and emphasise the need to use detailed 'real world' data collected as a joint effort between national population-based cancer registries and clinical registries, and routine health care data. By merging all this information, we should be able to conduct clinical outcomes research in an efficient way and to provide results that are relevant to all patient groups affected by the selected cancer type, rather than just those eligible for entry into clinical trials.

Herein, we describe the barriers in the development of evidence-based medicine and the possibilities of strong development in research and clinical investigations by using rare cancer examples. These illustrations and modalities to develop research in rare situations should be considered for all type of cancers since the concept of rare cancer is evolving – the genetics/unique biological features of a tumour may reveal a rare subtype of an otherwise 'common cancer' (i.e. lung, breast, colon).

The following examples illustrate the heterogeneity of rare cancers, as well as the different clinical contexts, incidence rates (Figure 1, Table 1) and the latest therapeutic progress:

- Childhood cancers: the care of children with cancer is based on decades of collaborative clinical trials, organised on a national or international level. All childhood cancers are currently defined as rare. This has led to long-term overall survival rates in excess of 80% in high income countries, but with variations ranging from 5% to >95% according to individual subtypes of childhood cancer [5]. Therein lies the challenge – further optimisation of therapy for those with a 'very good' prognosis on current standard therapies is increasingly difficult as the expected number of events reduces and sophisticated risk-stratification yields smaller subgroups [6].
- Haematological malignancies in adults (20% of incident rare adult cancer): a greater understanding of their biology and hence how they may be treated. Two rare leukaemia types (acute promyelocytic leukaemia and chronic myeloid leukaemia), which have poor prognosis when treated with cytotoxic chemotherapy, are now considered to have very favourable prognosis with targeted therapies [7].
- Sarcoma (5% of incident rare adult cancer): sarcomas can be classified into two

broad categories soft tissue sarcomas (STS), and sarcomas of the bone. They account for only ~1% of all adult solid malignant tumours, yet represent more than 70 distinct tumour subtypes. Obtaining the correct diagnosis of specific subtypes of sarcoma is becoming increasingly important in delivering optimal medical care [8].

2 Uncertainty in the diagnosis, in the treatment: ‘intrinsic lack or defect in evidence’: limited patients and scarce high-level evidence literature

For most rare cancers, research to identify causes (aetiology) or to develop strategies for prevention or early detection is limited or non-existent [9]. Even when a specific molecular defect underlying a particular rare cancer is discovered, it can still be challenging to get a quick and certain diagnosis. **Finally, standards of care derived from RCTs are not available for the majority of them and treatment options for the patient could be less effective, partly due to the non-availability of high grade evidence studies.**

Several studies have reported the frequency of histological diagnostic inaccuracies in rare tumours. For sarcomas, the study carried out by the Conticanet network on three European regions quantified the problem. This study was carried out on a complete series of tumours, collected through regional networks. All the tumours diagnosed and tumours suspected of corresponding to a sarcoma were reviewed by a panel of national and international expert pathologists. Significantly, the centralised review corrected diagnostic inaccuracies in a significant proportion of cases. Fifty per cent of the cases for which the first pathologist was uncertain about the diagnosis and requested a second opinion, were assigned a classification. Discrepancies were related to benign versus malignant, diagnosis of carcinoma versus sarcoma, and the incorrect diagnosis of histological subtypes and grade [10]. Requests by the primary pathologist involved about 30% of patients. Of course, these diagnostic parameters have a major influence not only on the subsequent therapeutic management but also on the interpretation of clinical research and the evidence base. Indeed, without centralised review, up to 28% of patients with rare tumours included in clinical trials may be misclassified, as published evidence on this in sarcoma and lymphomas ([11–13]). The centralised review of the diagnosis, implemented in clinical trials of rare tumours in many groups, including the European Organization for Research and Treatment of Cancer (EORTC), is therefore an essential tool for obtaining quality data. Another consequence of this misclassification concerns etiological research in case-control studies (the most frequent design in rare cancers), and underlines the need to carry out new studies with a correct inclusion of cases after central review.

The particularities for rare cancer impact directly on the patients for whom no ‘standard of care’ treatment exists, and so the first objective should be to devise an optimal treatment plan.

In theory, the same rules should apply for the definition of standard treatments in both rare and frequent tumours. For rare tumours, therapeutic standards have often been implemented from studies without a control arm or from RCTs whose small patient numbers mean they are underpowered to detect any differences [14]. The therapeutic standards are thus based on more 'fragile' criteria. It is not uncommon that no therapeutic standard is available in the absence of previous clinical studies.

Under these conditions, therapeutic 'standards' are applied to the patient with a significant amount of uncertainty. The definition of standard treatment against which to compare the new treatment in a clinical trial is confronted with the same level of uncertainty, and, often, this standard arises from expert consensus whose scientific basis remains low. Even so, international and national societies have developed, diffused and updated clinical practice guidelines (CPGs) for rare tumours: from diagnosis to care practices and treatment [15,16]. However, poor or moderate adherence with these clinical practice guidelines is frequently reported. For example, in STS, we observed low adherence regarding three fundamental elements in care management [10]: receiving histological diagnosis before surgery, adequacy of histological diagnosis and multidisciplinary discussion before surgery.

As a results of accumulation of defects in the detection, diagnosis and treatment, recent data revealed worse survival for rare cancers with differential situations for type of cancer, for age of patients or for stage at presentation [17]. EURO CARE has published substantial regional differences across EU countries in survival from rare cancers for which there are no effective treatments (e.g. mesothelioma), suggesting variations in the quality of diagnosis and follow-up. For rare cancers that respond well to treatment, differences in regional survival are possibly attributable to variations in treatment quality and availability or to cancer awareness in the population. Geographical variation in survival for these cancers might therefore reflect differences in the use of effective treatment protocols.

In response to questionable access and care management for rare tumours, several countries implemented specific actions according to specific rare cancers in national plans, e.g. threshold activity volumes for care management, secondary histological review and access to specialised national multidisciplinary teams or reference centres. Moreover, in some countries, rare cancer patient advocacy groups, such as Cancer 52 in the UK [18] and specific rare tumour groups and ECPC together with the RARECAREnet in the EU [19,20] have made efforts to improve patient information and access to services.

In March 2017, the European Union approved the launch of the first tranche of European Reference Networks (ERNs), part of a broader cross-border healthcare initiative to provide access to highly specialised diagnosis, treatment and advice for the 30 million citizens estimated to be affected by rare or low prevalence complex diseases. Three European Reference Networks cover all childhood cancers (ERN PAEDCAN) [21], all rare adult solid

tumours (ERN-EURACAN) [22] and haematological malignancies (EUROBLOODNET) [23]. These aims to facilitate management of rare cancers by provision of 'virtual' multidisciplinary tumour boards (MTB), clinical guidelines, training and stimulating research and innovation (Figure 2). In general, it is anticipated that the knowledge travels rather than the patient, though cross-border movement may be needed for a patient to access a particular complex therapy that requires concentrated expertise.

3 Why is it essential to give priority to evidence-based care in rare cancer?

3.1 The biology of rare cancer

Although most basic research and clinical trials, at least in adults, have historically focused on common cancers, it can be argued that a disproportionately large amount of understanding of cancer biology has come from the study of rare cancers [24]. The relatively recent and remarkable example of adult chronic myelogenous leukaemia has demonstrated how such biological knowledge can dramatically shorten the time taken from discovery of a targeted therapy to its use as standard of care [7].

Large-scale efforts, including the International Cancer Genome Consortium, the Cancer Genome Atlas, and the Cancer Genome Project, have catalogued genomic alterations in different cancer types and have provided the impetus for systematically incorporating molecular subclassification of cancer into histopathology and recognition of the common drivers of some rare entities

Rare cancers show some of the highest response rates, likely due to the identification of oncogenic drivers with little inter-patient variability [25]. As such, the imatinib story is remarkable. Imatinib is best known for revolutionising the outcome of chronic myelogenous leukaemia (CML), a rare illness whose hallmark is an aberrantly activated BCR-ABL fusion protein. Imatinib was designed to target this aberrant protein. Because of imatinib (as well as second generation BCR-ABL inhibitors), the median survival of patients with CML has increased from about five to more than twenty years. Imatinib is also effective in and approved for gastrointestinal stromal tumours, an uncommon type of neoplasm that usually carries a KIT mutation targetable by imatinib, and which was notoriously refractory to chemotherapy.

Advances in cancer biology and genomic technology have not only led to the definition of various molecularly defined rare cancers, but also are reshaping the focus and practice of drug development [24]. Molecular alterations identified in rare cancers can help with diagnosis and represent potential treatment targets. However, diagnostic entities are evolving rapidly, making consistent case identification difficult and confounding attempts at systematic data collection. Molecular sub-classification can be further extended to include host factors such as expression of immune markers and cell infiltrates, which are of great

interest because of the emergence of immunotherapies [26].

In this context, it makes sense to develop, at national and international levels, molecular analyses on each rare cancer at the time of first diagnosis, in order for the best molecular match to guide a management plan. Likely prognosis and the most appropriate management and treatment may be better estimated than from our current anatomical and histological characterisation.

3.2 From national cancer plans, international research collaborations and community commitment

In numerous countries, national cancer plans have encouraged collaborative research and guided the development of optimal care management for common cancers and rare cancers. Several national initiatives to organise care and research for rare cancers are currently operational, some for several years. As recently underlined [17], in France, a specific plan of action is dedicated to the development of specialized expert centres for rare cancer patients, labelled 'reference centres for rare cancers'. This organisation implements the creation of a network of regional centres for rare adult cancers referred to as national 'expert centres' (reference centres). Although some national reference networks have very strong coverage, **like in France** [27], with over 70% of new patients discussed at a referral MTB or recorded in the corresponding databases, other have still insufficient coverage. The organisation for rare cancers with discussion of files at a referral MTB and/or recording of cases in national databases act as a lever for encouraging clinical trials and facilitating access to innovative treatments. In clinical research, the International Rare Cancers Initiative (IRCI) was formed in 2011 as a partnership between several national organisations to stimulate and facilitate the development of international clinical trials for patients with rare cancers [28–30]. The IRCI focuses on interventional (usually randomised) clinical trials aiming to improve outcomes for patients. The objective of this initiative has been to facilitate the development of definitive treatment clinical trials for cancers that otherwise could not be addressed within individual regional catchments. Specifically, the initiative has focused on cancers with an incidence of less than 3 per 100,000 persons/year, for which no other international forum performs a coordinating function and for which treatment is not currently guided by strong evidence. Coordinated approaches for cooperative clinical trials therefore need to be improved in 2017. **Finally, the role of patient associations** in this process is important and needs strengthening, particularly the role in guiding the patient to reference centres and clinical trials. Patient advocacy groups have additional means of communicating with patients and their families, and through these communications they have the potential to speed dissemination of information on emerging cancer therapies. New social media platforms now enable virtual patient aggregation on an larger scale as showed by Katz [31]. Face-to-face

groups exist for many rare cancer types and could become a tool to accelerate clinical trials enrolment in these rare cancers.

The above ambitions are already being realised in the pilot European Expert Paediatric Oncology Reference Network for Diagnostics and Treatment, ExPO-r-Net [32], which was the forerunner for the development of the PaedCan ERN [21]. This pilot project established 'roadmaps' for four indications (retinoblastoma, liver tumours, renal tumours and very rare childhood cancers) that identified referral 'hubs' to provide highly specialised care and coordinated multidisciplinary advice for these rare cancer types that individually affect less than 1 per 100,000 young people per year.

4 Potential sources of relevant data for evidence-based practice and medical decisions

Studies of rare tumours present many challenges. Funding is limited; the pharmaceutical industry initially had little incentive to develop drugs for rare cancers; patient accrual to trials is frequently prolonged; there is no consensus about the most efficient clinical trial design methodology; and national regulatory requirements currently significantly impair the ability to conduct international trials. Rare cancer Europe recommendations [33] proposed several elements for improving clinical studies in rare cancers. In their basic considerations, they saw that the possible innovative solutions in designing trials may imply a price to pay in terms of higher uncertainty. Others underlined that all available data should be considered when establishing the efficacy of an intervention because data generated in rare cancer trials are generally less robust than those in studies of uncommon cancer [1].

4.1 'Traditional' evidence-based data

4.1.1 From RCTs

The term medicine-based evidence was coined to describe a rigorous and patient-centred approach to evaluate data. It recognises that RCTs may not always yield higher-quality evidence than observational studies, and/or provides high-quality evidence where RCT data are lacking. Evidence synthesis as trial meta-analyses are also less frequent in the context of scarce data. Thus, defining treatment standards for rare cancers (or rare subsets of cancers) remains a challenge. If a RCT cannot be finished how does one establish practice-changing results? [34].

Although single arm, non-randomised trials may be feasible and even lead to drug approval in rare and molecularly defined cancers, as outcomes improve and more treatments become available, there will remain a need for RCTs to compare the emerging available drugs and strategies. At the present time, few rare cancers can be treated by therapies backed by evidence gathered from RCTs. To compare therapies for many of these cancers would require trials on an international scale to gain sufficient statistical power. Such international

clinical trials would accrue participants faster and offer lower collective administrative costs than would local or regional trials, and collaborating investigators could capitalise on shared infrastructure, centralised resources, and existing networks. Despite their potential advantages, however, financial, logistical, and regulatory challenges make international trials challenging [25]. In a methodological review of existing literature for rare diseases, two types of innovative methods were identified: methods minimising trial samples size and methods maximising the number of on-treatment participants; for observational data, propensity scores and self-controlled design were the more used in the literature[35]. More recently, Parmar and colleagues presented solutions for designing randomised trials when the ideal sample size is considered larger than the number of participants that can be recruited in a reasonable time frame [36]. This article provides a systematic approach to clinical trial design and outlines a framework that sets out a series of considerations for the team designing a trial. It encourages a period of constructive deliberation about the various design elements, each of which is carefully reconsidered. Other approaches with flexible rules have been described or recommended for rare cancers at international level [28,33,37,38] including the Bayesian design, optimized external control group, adaptive design, or seamless phase II/II trials. Considering the endpoints of main analyses, using composite or surrogate endpoints (even if not validated) is suitable for some, as are the continuous measures of outcome, which can have greater statistical power than an associated dichotomous variable (eg. change in tumour size versus tumour response) [33,35,37]. Besides the design and endpoints elements, it is crucial in rare cancer trials to minimise missing data and protocol violations [37].

Finally, the major problem with RCTs is the lack of external validity of the results due to the restricted selection of patients and the low rate of participation in clinical trials.

In paediatric cancers, all of which are individually rare, much progress in improving survival rates has been made over the last several decades, through multinational co-operative group trials. Whilst many RCTs have been conducted to optimise the use of currently available chemotherapy drugs, similar levels of improvement in survival have been achieved through single arm studies, particularly in the very rare cancers such as hepatoblastoma [6,39]. The cited review includes the perspective of parents and highlights the logistics and ethical issues involved in improving recruitment to clinical trials for cancer in children. In tumour types where individual countries or centres see only a handful of new cases a year, it often occurs that the local clinicians will use the standard arm of these trials as current 'best practice', without formally registering the patient or their data into the trial database. Whilst such practice is understandable, it has the disadvantage that rare patient data are not collected or usable for benchmarking. The value of having detailed information on patient and tumour demographics in a single international database can be demonstrated by a recent

comparative analysis of outcomes of children with Wilms tumour treated in the same clinical trial conducted in Germany and the UK and which recruited >90% of the incident patient population in both countries [40]. The combination of material for biological studies with clinical information has allowed interrogation of the reasons for a small survival difference between the two countries. The authors suggest that health service factors affecting early recognition of childhood cancer in primary care may be main reason for the observed differences.

4.1.2To real world data

Approaches to summarize available evidence should be encouraged for rare cancers [41]. They should enable taking account of all available knowledge, which is particularly critical when direct experimental evidence is scanty or of suboptimal methodological quality. The principles of systematic reviews should be followed. In addition to RCTs, also uncontrolled trials, as well as observational studies should be considered when summarising the available evidence. Observational data can provide essential elements to measure effectiveness, the translation of efficacy in the real world (real world evidence) [42,43]. The major problem of these data is the internal validity but not the external. By using some large observational databases (such as cancer registries and cohorts and clinical registries, i.e hospital registries), it is possible to have valuable informative data providing the natural history and treatments of rare cancers [37], as emphasised in a recent editorial [44], for the Metasarc database [45]. In this example, the results among STS patients showed that combination of chemotherapy improves survival in initial treatment. The hazard ratio in Judson and colleagues' RCT was identical, but the trial was not powered to affirm significant results [46]. On the other hand, as recently reported [8], useful data were available by using rare cancer network, which collects information for care management systematically. In a network of reference centres for sarcoma patients, the authors assessed the outcome of 26,883 patients discussed in these MTBs. The NetSarc database includes patient characteristics, treatment and diagnostic procedures, as well as survival and progression rates. Relapse rates in this large real-life series of 26,883 sarcoma patients of the NETSARC network were higher than those in previous reports. Presentation in an MTB prior to first treatment is a major parameter associated with lower rates of relapse. It is essential to consider these results when we compare the results in an international STS trial, because the consequences are that STS patients managed in reference centres have a lower risk of relapse before the inclusion in any clinical trial.

The European Commission's initiative to establish Expert Reference Networks (ERNs) to improve care for patients with rare or low prevalence, complex diseases, will also stimulate clinical and translational research into rare cancers. The three ERNs that encompass haematological and solid tumours in adults and children are in the process of establishing

virtual tumour boards for cross-border discussion of rare and complex cases that will also build libraries of knowledge about treatment choices and outcomes and facilitate the necessary multinational level of research.

Statistical methods such as the well-known propensity score allow taking into consideration the selection bias in the comparison of different therapeutic strategies in observational studies [47]. Recently, Nussbaum et al used this approach in the US cancer national database to study the effect of radiotherapy on overall survival in patients with retroperitoneal sarcoma [48]. This question is presently addressed in a European Organisation for Research and Treatment in Cancer randomised study, comparing surgery alone, as the present standard of care, with preoperative external beam radiotherapy and surgery, as the experimental arm. For others, the retrospective, multi-institutional nature of this study without restriction to high-volume centres and with its acknowledged paucity of data about key aspects of sarcoma care such as surgical technique and margin status, is a massive limitation to the results [49].

As underlined by Billingham and colleagues [37], the next generation of large observational databases is likely to arise from the evolution of electronic patient records and should overcome the diversity of platforms between medical institutions. The final step will be to incorporate the data from research units (-omics data) and from administrative institutions (cost data).

4.1.3 And the right and big place of population-based cancer registries and international collaborations

The European Union Joint Action on Rare Cancers [50] aims to improve knowledge and quality of care for patients with rare cancers. This initiative includes collaborative research, guideline development and access to appropriate expertise that will be implemented through the relevant European Reference Networks. A major focus is to improve epidemiological surveillance of rare cancer entities and to enable population-based cancer registries to explore methods for efficient linkage to the increasingly rich datasets that exist in routine health care data, including molecular risk stratifications. Harmonised definitions of new data items and closer working with translational researchers to link patient outcomes to genomic information will be developed to enable the ERNs to establish the 'real world' benefits of more personalised, 'precision medicine' therapeutic approaches in rare cancers. A complementary approach to improving outcomes is illustrated by the International Cancer Benchmarking Partnership [51]. Here, eight countries collaborate to generate comparable data on public and patient awareness and diagnostic pathways from primary care to understand variation in survival, including outcome by tumour stage [52]. Whilst currently focussed on common cancers, this approach to benchmarking metrics that cover the entire patient pathway, including treatment and expected [52–54]

From rare data to potential big data of rare tumours

Biomedical big data is characterised by its size, variety of data sources and complexity. Today, the most popular definition of Big Data is the 5Vs, which are Volume, Velocity, Variety, Verification/Veracity, and Value. Big Data methods allow researchers to maximise the potential of existing data and enable new directions for research [55]. In oncology, the main challenge is for investigators to harness the power of Big Data to advance basic and clinical research whose goal is to discover novel ways to increase cancer cure rates and/or decrease treatment toxicity [56]. Big Data analytics covers collection, manipulation, and analyses of massive, diverse data sets that contain a variety of data types including genomic data and EHRs to reveal hidden patterns, cryptic correlations, and other intuitions on a Big Data infrastructure.

The generation of big data in cancer comes mainly from the high-throughput technologies used to study the 'omics sciences. Genomics concern genome sequence, structures, mutations, repeat contents and evolution. The next-generation sequencing (NGS), or high-throughput sequencing, is a technology that allows DNA and RNA sequencing much faster and cheaper than previously used methods.

To take medicine-based evidence in oncology to the next level, larger, comprehensive registries and big data are needed and will require appropriate funding.

The extraordinary potential of NGS technology makes it possible for rare cancer types to be divided into molecular 'subsets' for more accurate study. This may, paradoxically, reduce the ~200 rare cancer subtypes identified by RARECARE, to a more manageable number of 'molecular' groupings, providing some context as to prognosis and treatment direction for those patients for whom we currently have little in the way of evidence-based guidance. Many common cancers types may also become 'rare' by molecular association, as has been described above for molecular subsets of melanoma and lung cancer. Is genomics making rare cancers common and common cancers rare? (Sarcoma versus Colorectal cancer).

There is currently a huge volume of data sets that group information on cancer genome, transcriptome, clinical data and more. The challenge of big data in cancer is to integrate all this diversity of data collections into a unique platform. The possibility of harnessing information from all the accumulated data leads to an improvement in cancer patient treatment and outcome. A remarkable example is the big-data project CancerLinQ, from ASCO, that will allow patients and physicians to share information about treatments and outcomes [57]. It creates a continuous cycle of learning that begins and ends with the patient. One very important point is that patients are anonymised in databases to protect their identity. Patients and their doctors can contribute and gain from the accumulated information present in the database. For example, a type of cancer with a particular genetic

mutation is found to develop resistance to a targeted therapy. The shared information can prevent the physician from using the same drug in another patient with the same mutation.

4.2 Current challenges to develop evidence based medicine in rare cancer

RCTs are difficult to activate, and even more difficult to successfully complete, in rare or molecularly-defined cancers. Many have proposed different approaches for the future. For molecularly-targeted therapies, clinical trials based on cognate genomic aberrations should be the preferred pathway. For multimodality organ site-specific therapies, including surgery, medical therapies, and/or radiation therapy, open-ended single arm trials, or carefully performed prospective registry studies, are most likely to lead to new treatment standards. The challenge for cancer research is to better explore all data sets from tumour biology and clinical information about patients. This necessitates of an agreement on how these data are represented and accessed. Without such standards, the ability of both researchers and clinicians to make the best use of these data is severely limited.

4.2.1 Molecular challenges for trial design: basket and umbrella design and other platforms protocols

The clinical research community has generated novel trial designs including multiphase, multigroup adaptive, umbrella and basket trials with several defined cohorts for the testing of targeted therapies [58,59]. Such trials increase efficiency by providing opportunities to assess several drugs simultaneously within cohorts of patients with histologically and molecularly defined cancers. The cohorts can be modified with time depending on the drugs' activities. Regulators and funders of clinical trials seem to be open to novel trial designs to speed up assessment and time to approval, and willing to accept post-marketing assessments to generate additional safety and effectiveness data.

The 'basket trial' approach seems appropriate to develop drugs in rare cancers. Indeed, based on the observation that rare cancers show some of the highest response rates to molecularly targeted therapies, a more efficient way to study rare cancers is to use molecularly defined, histology-diagnostic trials designated as 'basket' studies. Such trials might include patients with a wide variety of histologies. An extension of the basket trial is the simultaneous study of multiple drugs/multiple histological subtypes, such as the one proposed in the Molecular Analysis for Therapy Choice (NCI-MATCH, NCI-COG) project for adults and children [60]. NCI-MATCH will prospectively enrol about three thousand patients who have progressed on at least one line of standard therapy. Patients will undergo a research-related biopsy for molecular tumour analysis, which will be performed in specified reference laboratories. At least twenty-five percent of the screened population will consist of patients with histologically rare cancers; genetic variants of common cancers do not count towards this number. The trial aims at developing a longitudinal cohort of patients with

multiple histologies who can be matched to one or more embedded single arm drug trials. Each trial will have dual objective response rate and progression free survival as the primary endpoints.

4.2.2 Data sharing, data harmonisation and standardization

Despite much progress, genomic and clinical data are still generally collected and studied in silos: by disease, by institution, and by country. Regulatory data-privacy requirements do not seamlessly lend themselves to the secure sharing of data within and across institutions and countries. Tools and analytical methods are non-standardised and incompatible, and the data are often stored in incompatible file formats. The volume of existing data is enormous and growing very fast. It is necessary to be able to exploit these existing data, asking complex questions to identify new knowledge [56]. Data standards in cancer research have evolved considerably in the past decade. Rapid developments in tumour classification and drug discovery are now overtaking the rate at which they are adopted into traditional vocabularies such as the International Classification of Diseases for Oncology. The evolution of clinical data standards has been taxonomised by medical informaticians from as early as the 1990s, covering a broad range of semantic and syntactic transitions. As databases rely on the accurate classification of data, changes have substantial effects in the way databases are modelled and structured, and subsequently in how they are queried. Therefore, current databases have a strong need for cancer ontologies that can standardise data accurately. The standards for electronic data recording provide the basis for standard medical ontologies including the Systematised Nomenclature of Medicine Clinical Terms and the Health Level Seven International protocol as well as cancer-specific ontologies. To conform to these standards, a metadata model is typically constructed for the data set. The construction of metadata is an often laborious, costly and time-consuming step in database development and requires careful planning, evaluation of the research protocols involved in the study, consultation with end-users of the database, and in the case of multiinstitutional studies, examination of the legacy data models already in place in each institution [61]. Many of the proposed solutions have been generated through open-source initiatives and have been instrumental to the development of cancer databases. However, previous audits of commonly used cancer ontologies have shown that they are far from perfect.

4.2.3 The need of the large Biobanks

In the last 10 years, several countries have set up large scale cohorts of volunteers from specified age ranges drawn from the general public, to underpin research into disease susceptibilities and associated life-style factors. Blood samples and additional phenotypic assessment data are taken at the time of obtaining consent for lifetime follow up through health record linkage. An example is the UK BioBank, whose resources have supported 162 cancer related studies to date (out of 846 received) [62]. Whilst many of these are

epidemiological in nature, the increasing availability of potentially linkable data sources such as ongoing medical prescriptions, shopping and exercise habits, means that modifiable lifestyle factors could be interrogated in the future. The Rare Cancers Europe initiative [63] provides policy and educational recommendations for standardising diagnosis and treatment and advocates for new approaches to research, including adaptive clinical trial designs and collaborative biobanks for medical research as crucial to advancing the development of new treatments in rare cancers. These should be maintained by dedicated personnel in a centralised way, to realise good quality control, proper access, regulatory and ethical competence and harmonisation and standardisation. Currently, major obstacles have to do with data protection. While data confidentiality needs to be protected by putting in place all reasonable available means, the right of patients to donate their tissues for research, if they will, should be protected as well. The patient should be able to give a 'broad consent' for his/her data and tissues to be used for research purposes by the treating institutions, avoiding the need to re-consent whenever a new retrospective research is decided. Of course, proper ethics and scientific reviewing mechanisms for new researches should be in place.

5 Conclusion

Regardless of whether a disease is rare or common, it is important to consider the totality of the research evidence in order to determine true findings vs false findings. RCT are always desirable to answer the uncertainty related to a treatment efficacy but other ways can be developed as underlined in this work. Continued advances in accurate diagnosis of rare cancers and in clinical and translational research aim to improve treatment through a more biology-driven approach. There are several examples that illustrate the development of major innovative therapies using this approach. As it becomes progressively easier to collect vast amounts of disparate personal health- and population-related information on a global scale, the real challenge will be to curate, store, federate, integrate, share, mine, interpret, and transform these extensive heterogeneous data into scalable and reliable medically actionable resources. International collaboration is mandatory to reach these objectives and is ongoing through several initiatives for rare tumours. Even though methodological difficulties remain, the means required to overcome the obstacles are now visible.

Note Fig reference Steliarova-Foucher: [64]

6 References

- [1] Komatsubara KM, Carvajal RD. The promise and challenges of rare cancer research. *Lancet Oncol* 2016;17:136–8. doi:10.1016/S1470-2045(15)00485-4.
- [2] Gatta G, van der Zwan JM, Casali PG, Siesling S, Dei Tos AP, Kunkler I, et al. Rare cancers are not so rare: the rare cancer burden in Europe. *Eur J Cancer* 2011;47:2493–511. doi:10.1016/j.ejca.2011.08.008.
- [3] Gatta G, Capocaccia R, Botta L, Mallone S, De Angelis R, Ardanaz E, et al. Burden and centralised treatment in Europe of rare tumours: results of RARECAREnet-a population-based study. *Lancet Oncol* 2017;18:1022–39. doi:10.1016/S1470-2045(17)30445-X.
- [4] DeSantis CE, Kramer JL, Jemal A. The burden of rare cancers in the United States. *CA Cancer J Clin* 2017;67:261–72. doi:10.3322/caac.21400.
- [5] Gatta G, Botta L, Rossi S, Aareleid T, Bielska-Lasota M, Clavel J, et al. Childhood cancer survival in Europe 1999-2007: results of EURO CARE-5--a population-based study. *Lancet Oncol* 2014;15:35–47. doi:10.1016/S1470-2045(13)70548-5.
- [6] Pritchard-Jones K, Dixon-Woods M, Naafs-Wilstra M, Valsecchi MG. Improving recruitment to clinical trials for cancer in childhood. *Lancet Oncol* 2008;9:392–9. doi:10.1016/S1470-2045(08)70101-3.
- [7] Bachireddy P, Burkhardt UE, Rajasagi M, Wu CJ. Haematological malignancies: at the forefront of immunotherapeutic innovation. *Nat Rev Cancer* 2015;15:201–15. doi:10.1038/nrc3907.
- [8] Ray-Coquard I, Pujade Lauraine E, Le Cesne A, Pautier P, Vacher Lavenue MC, Trama A, et al. Improving treatment results with reference centres for rare cancers: where do we stand? *Eur J Cancer* 2017;77:90–8. doi:10.1016/j.ejca.2017.02.006.
- [9] Charbotel B, Fervers B, Droz JP. Occupational exposures in rare cancers: A critical review of the literature. *Crit Rev Oncol Hematol* 2014;90:99–134. doi:10.1016/j.critrevonc.2013.12.004.
- [10] Ray-Coquard I, Montesco MC, Coindre JM, Dei Tos AP, Lurkin A, Ranchère-Vince D, et al. Sarcoma: concordance between initial diagnosis and centralized expert review in a population-based study within three European regions. *Ann Oncol* 2012;23:2442–9. doi:10.1093/annonc/mdr610.
- [11] Proctor IE, McNamara C, Rodriguez-Justo M, Isaacson PG, Ramsay A. Importance of expert central review in the diagnosis of lymphoid malignancies in a regional cancer network. *J Clin Oncol* 2011;29:1431–5. doi:10.1200/JCO.2010.31.2223.
- [12] Wilkins BS. Pitfalls in lymphoma pathology: avoiding errors in diagnosis of lymphoid tissues. *J Clin Pathol* 2011;64:466–76. doi:10.1136/jcp.2010.080846.
- [13] Thway K, Wang J, Mubako T, Fisher C. Histopathological diagnostic discrepancies in soft tissue tumours referred to a specialist centre: reassessment in the era of ancillary molecular diagnosis. *Sarcoma* 2014;2014:686902. doi:10.1155/2014/686902.
- [14] Gaddipati H, Liu K, Pariser A, Pazdur R. Rare cancer trial design: lessons from FDA approvals. *Clin Cancer Res* 2012;18:5172–8. doi:10.1158/1078-0432.CCR-12-1135.
- [15] Hochhaus A, Saussele S, Rosti G, Mahon F-X, Janssen JJWM, Hjorth-Hansen H, et al. Chronic myeloid leukaemia: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2017;28:iv41-iv51. doi:10.1093/annonc/mdx219.

- [16] Moreau P, San Miguel J, Sonneveld P, Mateos MV, Zamagni E, Avet-Loiseau H, et al. Multiple myeloma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2017;28:iv52-iv61. doi:10.1093/annonc/mdx096.
- [17] Gatta G, Trama A, Capocaccia R. Epidemiology of rare cancers and inequalities in oncologic outcomes. *Eur J Surg Oncol* 2017. doi:10.1016/j.ejso.2017.08.018.
- [18] The common voice for less common cancers n.d. <http://www.cancer52.org.uk/> (accessed October 31, 2017).
- [19] Sarcoma Patients EuroNet n.d. <http://www.sarcoma-patients.eu/en/> (accessed March 11, 2017).
- [20] Information Network on Rare Cancers 2017. <http://www.rarecarenet.eu/rarecarenet/index.php/patient-organisations> (accessed October 27, 2017).
- [21] ERN PaedCan – European Reference Network for Paediatric Oncology n.d. <http://paedcan.ern-net.eu/>.
- [22] European Reference Network Adult Cancers 2017. <http://euracan.ern-net.eu> (accessed November 11, 2017).
- [23] European Reference Network in Rare Hematological Diseases 2017. <https://www.eurobloodnet.eu> (accessed December 11, 2017).
- [24] Boyd N, Dancey JE, Gilks CB, Huntsman DG. Rare cancers: a sea of opportunity. *Lancet Oncol* 2016;17:e52–61. doi:10.1016/S1470-2045(15)00386-1.
- [25] Schott AF, Welch JJ, Verschraegen CF, Kurzrock R. The National Clinical Trials Network: Conducting Successful Clinical Trials of New Therapies for Rare Cancers. *Semin Oncol* 2015;42:731–9. doi:10.1053/j.seminoncol.2015.07.010.
- [26] Sharifnia T, Hong AL, Painter CA, Boehm JS. Emerging Opportunities for Target Discovery in Rare Cancers. *Cell Chem Biol* 2017;24:1075–91. doi:10.1016/j.chembiol.2017.08.002.
- [27] Pasquali S, Bonvalot S, Tzanis D, Casali PG, Trama A, Gronchi A, et al. Treatment challenges in and outside a network setting: Soft tissue sarcomas. *Eur J Surg Oncol* 2017. doi:10.1016/j.ejso.2017.09.015.
- [28] Bogaerts J, Sydes MR, Keat N, McConnell A, Benson A, Ho A, et al. Clinical trial designs for rare diseases: studies developed and discussed by the International Rare Cancers Initiative. *Eur J Cancer* 2015;51:271–81. doi:10.1016/j.ejca.2014.10.027.
- [29] Keat N, Law K, Seymour M, Welch J, Trimble T, Lascombe D, et al. International rare cancers initiative. *Lancet Oncol* 2013;14:109–10. doi:10.1016/S1470-2045(12)70570-3.
- [30] International Rare Cancers Initiative n.d. <http://www.irci.info/>.
- [31] Katz MS, Utengen A, Anderson PF, Thompson MA, Attai DJ, Johnston C, et al. Disease-Specific Hashtags for Online Communication About Cancer Care. *JAMA Oncol* 2016;2:392–4. doi:10.1001/jamaoncol.2015.3960.
- [32] European Expert Paediatric Oncology Reference Network for Diagnostics and Treatment n.d. <http://www.expornet.eu/> (accessed October 30, 2017).
- [33] Casali PG, Bruzzi P, Bogaerts J, Blay J-Y, Rare Cancers Europe (RCE) Consensus Panel. Rare Cancers Europe (RCE) methodological recommendations for clinical studies in rare cancers: a European consensus position paper. *Ann Oncol* 2015;26:300–6. doi:10.1093/annonc/mdu459.
- [34] Brade AM, Dawson LA. To RCT or Not to RCT: How to Change Practice for Rare Cancers? *J Clin Oncol Off J Am Soc Clin Oncol* 2016;34:203–4. doi:10.1200/JCO.2015.63.8767.

- [35] Gagne JJ, Thompson L, O’Keefe K, Kesselheim AS. Innovative research methods for studying treatments for rare diseases: methodological review. *BMJ* 2014;349:g6802.
- [36] Parmar MKB, Sydes MR, Morris TP. How do you design randomised trials for smaller populations? A framework. *BMC Med* 2016;14:183. doi:10.1186/s12916-016-0722-3.
- [37] Billingham L, Malottki K, Steven N. Research methods to change clinical practice for patients with rare cancers. *Lancet Oncol* 2016;17:e70–80. doi:10.1016/S1470-2045(15)00396-4.
- [38] Hilgers R-D, König F, Molenberghs G, Senn S. Design and analysis of clinical trials for small rare disease populations. *J Rare Dis Res Treat* 2016;1:53–60.
- [39] Perilongo G, Brown J, Shafford E, Brock P, De Camargo B, Keeling JW, et al. Hepatoblastoma presenting with lung metastases: treatment results of the first cooperative, prospective study of the International Society of Paediatric Oncology on childhood liver tumors. *Cancer* 2000;89:1845–53.
- [40] Pritchard-Jones K, Graf N, van Tinteren H, Craft A. Evidence for a delay in diagnosis of Wilms’ tumour in the UK compared with Germany: implications for primary care for children. *Arch Dis Child* 2016;101:417–20. doi:10.1136/archdischild-2015-309212.
- [41] Behera M, Kumar A, Soares HP, Sokol L, Djulbegovic B. Evidence-based medicine for rare diseases: implications for data interpretation and clinical trial design. *Cancer Control* 2007;14:160–6.
- [42] Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-World Evidence - What Is It and What Can It Tell Us? *N Engl J Med* 2016;375:2293–7. doi:10.1056/NEJMSb1609216.
- [43] Schilsky RL. Finding the Evidence in Real-World Evidence: Moving from Data to Information to Knowledge. *J Am Coll Surg* 2017;224:1–7. doi:10.1016/j.jamcollsurg.2016.10.025.
- [44] Benjamin RS. Observational studies: goldmines of information on rare diseases. *BMC Med* 2017;15:100. doi:10.1186/s12916-017-0868-7.
- [45] Savina M, Le Cesne A, Blay J-Y, Ray-Coquard I, Mir O, Toulmonde M, et al. Patterns of care and outcomes of patients with METAstatic soft tissue SARComa in a real-life setting: the METASARC observational study. *BMC Med* 2017;15:78. doi:10.1186/s12916-017-0831-7.
- [46] Judson I, Verweij J, Gelderblom H, Hartmann JT, Schöffski P, Blay J-Y, et al. Doxorubicin alone versus intensified doxorubicin plus ifosfamide for first-line treatment of advanced or metastatic soft-tissue sarcoma: a randomised controlled phase 3 trial. *Lancet Oncol* 2014;15:415–23. doi:10.1016/S1470-2045(14)70063-4.
- [47] Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *Am J Epidemiol* 2003;158:280–7.
- [48] Nussbaum DP, Rushing CN, Lane WO, Cardona DM, Kirsch DG, Peterson BL, et al. Preoperative or postoperative radiotherapy versus surgery alone for retroperitoneal sarcoma: a case-control, propensity score-matched analysis of a nationwide clinical oncology database. *Lancet Oncol* 2016;17:966–75. doi:10.1016/S1470-2045(16)30050-X.
- [49] Gronchi A, Haas RL, Bonvalot S. Cancer registries and randomised clinical trials in rare tumours: At the two extremes of daily clinical practice. *Eur J Cancer* 2016;64:113–5. doi:10.1016/j.ejca.2016.05.025.

- [50] JointActionRareCancers 2017. <http://jointactionrarecancers.eu/> (accessed November 14, 2017).
- [51] International Cancer Benchmarking Partnership n.d. www.icbp.org.uk (accessed December 11, 2017).
- [52] Butler J, Foot C, Bomb M, Hiom S, Coleman M, Bryant H, et al. The International Cancer Benchmarking Partnership: an international collaboration to inform cancer policy in Australia, Canada, Denmark, Norway, Sweden and the United Kingdom. *Health Policy Amst Neth* 2013;112:148–55. doi:10.1016/j.healthpol.2013.03.021.
- [53] Weller D, Vedsted P, Anandan C, Zalounina A, Fourkala EO, Desai R, et al. An investigation of routes to cancer diagnosis in 10 international jurisdictions, as part of the International Cancer Benchmarking Partnership: survey development and implementation. *BMJ Open* 2016;6:e009641. doi:10.1136/bmjopen-2015-009641.
- [54] Rose PW, Rubin G, Perera-Salazar R, Almberg SS, Barisic A, Dawes M, et al. Explaining variation in cancer survival between 11 jurisdictions in the International Cancer Benchmarking Partnership: a primary care vignette survey. *BMJ Open* 2015;5:e007212. doi:10.1136/bmjopen-2014-007212.
- [55] Ehrenstein V, Nielsen H, Pedersen AB, Johnsen SP, Pedersen L. Clinical epidemiology in the era of big data: new opportunities, familiar challenges. *Clin Epidemiol* 2017;9:245–50. doi:10.2147/CLEP.S129779.
- [56] Barbosa C. Challenges with Big Data in Oncology. *J Orthop Oncol* 2016;2.
- [57] Sledge GW, Miller RS, Hauser R. CancerLinQ and the future of cancer care. *Am Soc Clin Oncol Educ Book Am Soc Clin Oncol Meet* 2013:430–4. doi:10.1200/EdBook_AM.2013.33.430.
- [58] Renfro LA, Sargent DJ. Statistical controversies in clinical research: basket trials, umbrella trials, and other master protocols: a review and examples. *Ann Oncol* 2017;28:34–43. doi:10.1093/annonc/mdw413.
- [59] Redig AJ, Jänne PA. Basket trials and the evolution of clinical trial design in an era of genomic medicine. *J Clin Oncol* 2015;33:975–7. doi:10.1200/JCO.2014.59.8433.
- [60] National Cancer Institute: Molecular Analysis for Therapy Choice (NCIMATCH) n.d. <https://www.cancer.gov/about-cancer/treatment/clinical-trials/nci-supported/nci-match> (accessed October 30, 2017).
- [61] Noor AM, Holmberg L, Gillett C, Grigoriadis A. Big Data: the challenge for small research groups in the era of cancer genomics. *Br J Cancer* 2015;113:1405–12. doi:10.1038/bjc.2015.341.
- [62] Biobank UK 2016. <http://www.ukbiobank.ac.uk/approved-research/> (accessed October 24, 2017).
- [63] Rare Cancers Europe 2017. <http://www.rarecancerseurope.org/> (accessed February 11, 2017).
- [64] Steliarova-Foucher E, Colombet M, Ries LAG, Moreno F, Dolya A, Bray F, et al. International incidence of childhood cancer, 2001-10: a population-based registry study. *Lancet Oncol* 2017;18:719–31. doi:10.1016/S1470-2045(17)30186-9.

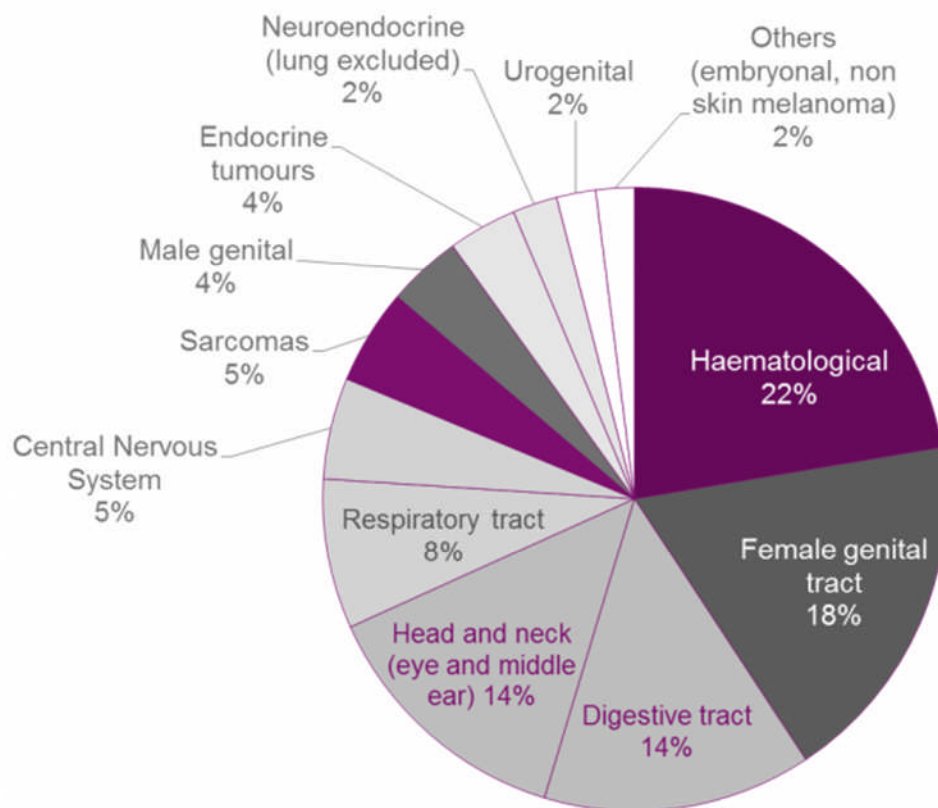


Figure 1. Distribution of families of rare cancers in adults people. <http://www.rarecancerseurope.org/About-Rare-Cancers/Families-and-List-of-Rare-Cancers> (graph courtesy of the RARECARE project)

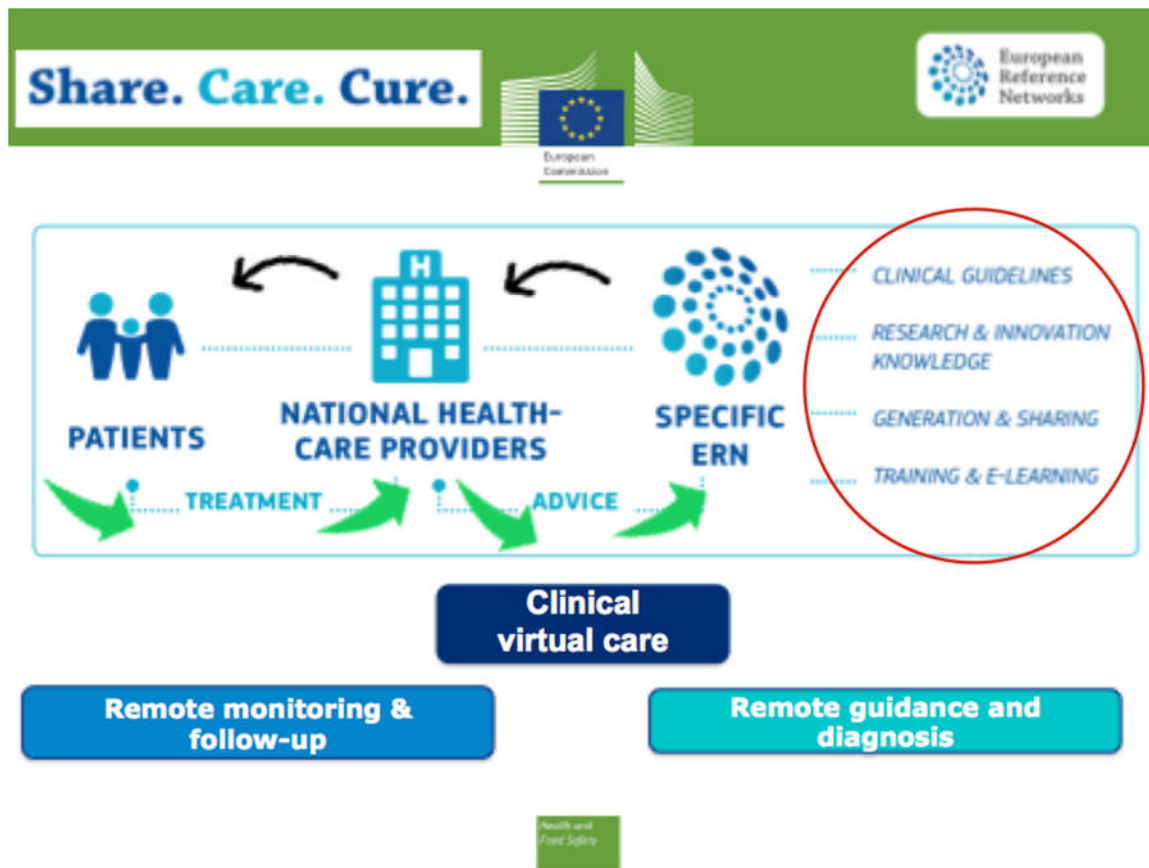


Figure 2. European Reference Networks (ERNs) Model: ERN are virtual networks involving healthcare providers across Europe. They aim to tackle complex or rare diseases and conditions that require highly specialised treatment and concentrated knowledge and resources (https://ec.europa.eu/health/ern_en). Courtesy of Enrique Terol, scientific officer, European Commission.

Table 1. Crude incidence and relative survival rates for some rare adults cancers in Europe (Gatta 2017 Lancet oncol)

	Crude incidence/ 100 000 (EU)	5 year-relative Survival
Acute myeloid leukaemia	3.50	17.5
Chronic myeloid leukemia	1.12	54.9
Soft tissue sarcoma	4.71	56.7
Malignant mesothelioma	2.14	5.3
Bone Sarcoma	0.85	58.6
All rare adult cancers	115	48.5
<hr/>		
All childhood cancers	14.1*	77.9**

* figure quoted is age-standardised rate (world standard) for children aged 0-14yrs [60].

There are 12 major categories of childhood cancer, each with several sub-categories and within these, further subtypes, so each is individually. The equivalent estimated incidence rate in Europe from ref 5 is 13.4 per 100,000.

** figure quoted is 5yr observed survival (OS) for cases diagnosed in the period 2000-2007. OS in childhood is nearly equivalent to relative survival since competing risks of death are almost negligible [5].

