Estimating the total incidence of global childhood cancer: a simulation-based analysis

Zachary J Ward, Jennifer M Yeh, Nickhill Bhakta, A Lindsay Frazier, Rifat Atun**

Summary

Background Accurate estimates of childhood cancer incidence are important for policy makers to inform priority setting and planning decisions. However, many countries do not have cancer registries that quantify the incidence of childhood cancer. Moreover, even when registries do exist, they might substantially underestimate the true incidence, since children with cancer might not be diagnosed. We therefore aimed to provide estimates of total childhood cancer incidence accounting for underdiagnosis.

Methods We developed a microsimulation model to simulate childhood cancer incidence for 200 countries and territories worldwide, taking into account trends in population growth and urbanicity, geographical variation in cancer incidence, and health system barriers to access and referral that contribute to underdiagnosis. To ensure model results were consistent with epidemiological data, we calibrated the model to publicly available cancer registry data using a Bayesian approach in which the observed data are fixed and the model parameters (cancer incidence and probabilities of health system access and referral) are random variables. We estimated the total incidence of childhood cancer (diagnosed and undiagnosed) in each country in 2015 and projected the number of cases from 2015 to 2030.

Findings Our model estimated that there were 397 000 (95% uncertainty interval [UI] 377 000–426 000) incident cases of childhood cancer worldwide in 2015, of which only 224 000 (95% UI 216 000–237 000) were diagnosed. This finding suggests that 43% (172 000 of 397 000) of childhood cancer cases were undiagnosed globally, with substantial variation by region, ranging from 3% in western Europe (120 of 4300) and North America (300 of 10 900) to 57% (43 000 of 76 000) in western Africa. In south Asia (including southeastern Asia and south-central Asia), the overall proportion of undiagnosed cases was estimated to be 49% (67 000 of 137 000). Taking into account population projections, we estimated that there will be 6·7 million (95% UI 6·3–7·2) cases of childhood cancer worldwide from 2015 to 2030. At current levels of health system performance, we estimated that 2·9 million (95% UI 2·7–3·3) cases of childhood cancer will be missed between 2015 and 2030.

Interpretation Childhood cancer is substantially underdiagnosed, especially in south Asia and sub-Saharan Africa (including western, eastern, and southern Africa). In addition to improving treatment for childhood cancer, health systems must be strengthened to accurately diagnose and effectively care for all children with cancer. As countries expand universal health coverage, these estimates of total incidence will hopefully help guide efforts to appropriately increase health system capacity to ensure access to effective childhood cancer care.

Funding Boston Children's Hospital, Dana-Farber Cancer Institute, Harvard T H Chan School of Public Health, Harvard Medical School, National Cancer Institute, SickKids, St Jude Children's Research Hospital, and Union for International Cancer Control.

Lancet Oncol 2019; 20: 483–93

Published Online February 26, 2019 http://dx.doi.org/10.1016/S1470-2045(18)30909-4

See Comment page 460

*Co-senior authors

Center for Health Decision Science (Z J Ward MPH) **and Department of Global Health and Population** (Prof R Atun FRCP), **Harvard T H Chan School of Public Health, Harvard University, Boston, MA, USA; Division of General Pediatrics, Boston Children's Hospital, Boston, MA, USA** (J M Yeh PhD); **Department of Pediatrics** (J M Yeh) **and Department of Global Health and Social Medicine** (Prof R Atun), **Harvard Medical School, Harvard University, Boston, MA, USA; Department of Global Pediatric Medicine, St Jude Children's Research Hospital, Memphis, TN, USA** (N Bhakta MD); **and Dana-Farber/Boston Children's Cancer and Blood Disorders Center, Boston, MA, USA** (A L Frazier MD)

Correspondence to: Mr Zachary J Ward, Center for Health Decision Science, Harvard T H Chan School of Public Health, Boston, MA 02115, USA **zward@hsph.harvard.edu**

Introduction

Childhood cancer—defined here as cancer in children aged 0–14 years—is a major cause of death in children worldwide.[1,2] More than 80% of diagnosed cases of childhood cancer occur in low-income and middle-income countries,[3,4] where access to diagnostics and treatment are often inadequate.[5] Accurate estimates of incidence are important for cancer control strategies, especially for countries with substantial population growth and those expanding universal health coverage, an important target for the Sustainable Development Goal (SDG) 3.[6] However, current estimates of cancer incidence,[1,2,7–9] based on reported data from population-based cancer registries, have severe data limitations related to both access and quality, especially in low-income and middle-income countries. Worldwide, about 60% of countries do not have quality population-based cancer registries, and those that do often cover only a small fraction of the population.[1] Indeed, only an estimated 11·4% of the world population aged 0–14 years was covered by cancer registries in 2000–10.[7] Where registries do exist, weak health systems in low-income and middle-income countries mean that many patients with cancer are not diagnosed and therefore not registered.[10] This underdiagnosis might be due to poor access to primary care (leading to an eventual death from the disease at home) or misdiagnosis due to inadequate diagnostics (eg, lymphoma misdiagnosed as tuberculosis).

Research in context

Evidence before this study
Until now, estimates of childhood cancer incidence have been based on unadjusted aggregated data from population-based cancer registries worldwide. The Global Burden of Disease Study 2016 provided detailed estimates of the number of incident childhood cancer cases for the year 2016, and the International Agency for Research on Cancer GLOBOCAN 2018 study provided estimates for the year 2018. We searched PubMed for studies on the incidence of global childhood cancer using the search terms "childhood cancer", "incidence", and "global" on Nov 12, 2018, without language or publication date restrictions. We found no other estimates of global childhood cancer incidence. Current estimates of global childhood cancer incidence are about 200 000 cases per year. However, these estimates do not adjust for underdiagnosis due to weaknesses in health systems.

Added value of this study
Health system barriers result in substantial underdiagnosis of childhood cancer cases in many countries. The true incidence of global childhood cancer is likely to be substantially higher than currently reported. We developed a novel simulation model of total childhood cancer incidence for 200 countries and territories worldwide that takes into account the effects of health system barriers on cancer diagnosis. We provided estimates of underdiagnosis by country and territory, and we estimated the total global incidence of childhood cancer.

Implications of all the available evidence
We estimated total global childhood cancer incidence to be close to 400 000 cases per year, suggesting that nearly one-in-two children with cancer are never diagnosed. These new estimates could help to guide the expansion of access to childhood cancer care in health systems that are expanding universal health coverage.

Although previous studies[7] acknowledge that underdiagnosis might contribute to low incidence rates of registry-reported childhood cancer in low-income and middle-income countries and beyond, no study has yet attempted to quantify its extent. We therefore aimed to use a simulation model to estimate country-specific childhood cancer incidence. By considering health system barriers that contribute to underdiagnosis in registry data, we aimed to provide new estimates of the total incidence of childhood cancer to inform health system policies for effective diagnosis and treatment of all children with cancer.

Methods

Study design and data sources

In this study, we developed the Global Childhood Cancer (GCC) microsimulation model (ie, an individual-level simulation model) to estimate childhood cancer incidence for 200 countries and territories worldwide, taking into account trends in population growth and urbanisation, geographical variation in cancer incidence, and health system barriers that contribute to underdiagnosis of childhood cancer. We used the model to estimate the total incidence of childhood cancer (both diagnosed and undiagnosed) and to estimate the total number of cases from 2015 to 2030 (the time period of the SDGs).[6]

See Online for appendix

Procedures

We developed a conceptual cancer diagnosis cascade (figure 1). Using this framework, we simulated children with cancer from incidence to diagnosis and registration. By modelling the process by which patients with cancer are identified and diagnosed, we leveraged available demographic, cancer incidence, and health systems data to estimate the effects of health system barriers on underdiagnosis of childhood cancer cases by calibrating our predicted rates of diagnosed cases to the reported incidence rates in country-specific cancer registries.

There are few known environmental risk factors for childhood cancer, and an underlying genetic predisposition is estimated to account for less than 10% of all childhood cancers (although no global population-based estimates of allele frequency variation exist).[11,12] We therefore assumed that populations with similar genetic composition (which, similar to GLOBOCAN,[8] we assume is based on geographical proximity) have similar rates of childhood cancer incidence, enabling us to use data from nearby countries with cancer registries to estimate incidence in countries without registry data. Nearby countries are also likely to share environmental exposures that might affect the incidence of some cancers.

Furthermore, by considering countries with similar genetic make-up and environmental exposures, we were able to exploit variability in health system performance to estimate the extent of underdiagnosis. For each country, we modelled the key health system barriers of access to primary care and appropriate referral to specialty care (appendix pp 37–41).

We synthesised country-specific data for demographics, cancer incidence, and health system variables from multiple sources to create a virtual population representative of global childhood cancer (table). We grouped countries into four income categories as defined by the World Bank (appendix p 2),[13] and 21 geographical regions as defined by the UN. We excluded areas not classified by the World Bank (appendix pp 2, 3). Our final model included 200 countries and territories.

We modelled population growth in each country using the UN probabilistic projections,[14] with annual projections[15] used to incorporate the distribution of individual ages and interpolate the 5-year probabilistic projections using linear interpolation (appendix p 3). We imputed projections based on regional trends for countries without UN projections. Specifically, we imputed projections of percentage change in population size based on the regional average. We also imputed uncertainty around these projections. By sampling population trajectories from the uncertainty intervals for each country, we accounted for the uncertainty of these projections, helping to guard against the potential error associated with any single projection of trends. We also modelled the urban or rural location of individuals based on the UN 2014 Urbanization Prospects.[16]

We obtained information about registry-reported cancer cases from the International Incidence of Childhood Cancer, volume III (IICC-3).[17] Registries from 77 countries were included in the model (appendix pp 4–7). We separately modelled each of the 48 cancer subcategories defined by the International Classification of Childhood Cancer.[19] For each diagnosis and age group (<1, 1–4, 5–9, and 10–14 years), we estimated hierarchical models of cancer incidence with four levels (global, continent, region, and country) weighted by the person-years of each registry (appendix pp 8–36). This approach allowed us to use all available registry data while maintaining geographical differences in reported cancer incidence. In particular, our hierarchical approach allowed us to cluster trends in cancer incidence at different geographical levels, with a flexible country-level term allowing us to model substantial heterogeneity within regions where appropriate, thus incorporating uncertainty around our guiding assumption that nearby countries have similar incidence rates.

We modelled access to primary care as the first step in the cancer diagnosis cascade. Given access to primary care, we assumed that patients must then be appropriately referred to specialty care (and successfully complete that referral) to receive an accurate cancer diagnosis. We assumed that all diagnosed patients were recorded in a cancer registry if one exists. To estimate the probabilities of access and referral, we leveraged information from proxy indicators. The indicators we chose are also included in the service coverage index[20] of universal health coverage and the WHO reference list of core health indicators,[21] suggesting good face validity. To model access, we selected indicators for access to primary prevention and care interventions: antenatal care coverage, vaccination coverage, and the WHO composite coverage indicator consisting of eight reproductive, maternal, newborn, and child health interventions (appendix pp 37, 38).

To model referral, the next step in the cascade, we selected indicators for receiving appropriate treatment for a given illness: suspected pneumonia referral and diarrhoeal treatment with oral rehydration salts or therapy (appendix pp 39, 40). Although diarrhoea can be treated at home, children must still receive appropriate health services to initiate treatment. Thus, these indicators can provide insight into the extent of health service engagement at each step of the cancer diagnosis cascade.

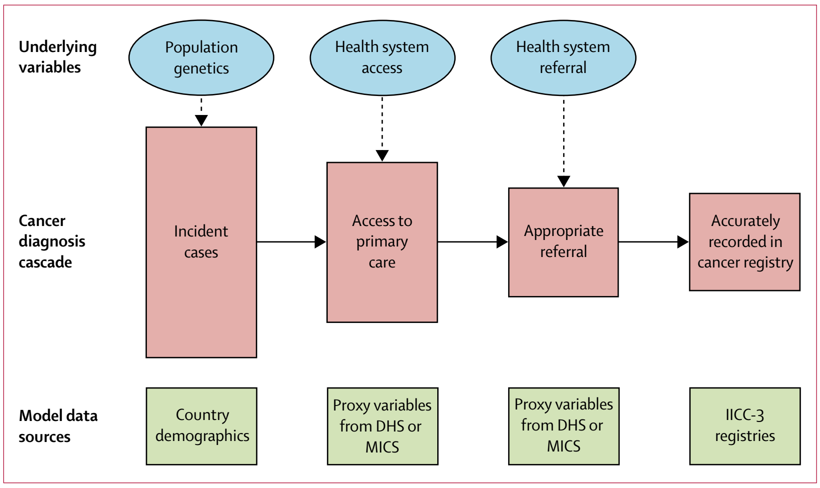

Figure 1: **Conceptual cancer diagnosis cascade**

DHS=Demographic Health Survey. IICC-3=International Incidence of Childhood Cancer, volume III. MICS=Multiple Indicator Cluster Survey.

We obtained the most recent data for each proxy indicator from the WHO Global Health Observatory data repository.[18] These country-specific indicators are based on Demographic Health Survey (DHS) or Multiple Indicator Cluster Survey (MICS) data, stratified by urban or rural location. Since most cancer registries are located in urban areas and therefore might not be nationally representative,[10] we modelled health system variables by urban or rural location to capture within-country variation in underdiagnosis of childhood cancers. To estimate prior probability distributions for access and referral based on these indicators, we developed a Bayesian hierarchical framework[22] with three levels (income, region, and country), allowing us to synthesise multiple indicators and estimate parameters for countries with no DHS or MICS data (appendix p 41).

Outcomes

For each country or territory, we modelled the effect of health system barriers on childhood cancer diagnosis and estimated the total incidence (ie, diagnosed and undiagnosed) of each International Classification of Childhood Cancer diagnosis by age group. We also projected the number of childhood cancer cases from 2015 to 2030, taking into account trends in population growth and urbanicity. We report the estimated mean and 95% uncertainty interval (UI; calculated as the 2·5 and 97·5 percentiles) of our simulation results.

Statistical analysis

For each age group in each country and territory, we simulated the cancer diagnosis cascade for each of the

	Data source	Number of model countries reported	Reference
Demographics			
Income group	World Bank 2016 income categories	200	World Bank 2016[13]
Population projections	Probabilistic Population Projections used for population estimates; Medium Fertility Scenario used for age structure	186	UN World Population Prospects: the 2015 revision[14,15]
Urban percentage	Proportion of population living in urban areas	200	UN 2014 Urbanization Prospects[16]
Cancer incidence			
Reported cancer cases	Global, continental, and regional estimates used as prior probability distributions; country-specific estimates used as calibration targets	77	International Incidence of Childhood Cancer-3 registries[17]
Health system variables			
Access (urban and rural)	Antenatal care coverage, at least four visits	86	DHS and MICS data, obtained from the WHO Global Health Observatory[18]
Access (urban and rural)	Composite coverage index	86	DHS and MICS data, obtained from the WHO Global Health Observatory[18]
Access (urban and rural)	Coverage of DTP3 vaccination	95	DHS and MICS data, obtained from the WHO Global Health Observatory[18]
Referral (urban and rural)	Children aged <5 years with pneumonia symptoms taken to a health facility	90	DHS and MICS data, obtained from the WHO Global Health Observatory[18]
Referral (urban and rural)	Children aged <5 years with diarrhoea receiving oral rehydration salts	99	DHS and MICS data, obtained from the WHO Global Health Observatory[18]
Referral (urban and rural)	Children aged <5 years with diarrhoea receiving oral rehydration therapy and continued feeding	99	DHS and MICS data, obtained from the WHO Global Health Observatory[18]

DHS=Demographic and Health Survey. DTP3=diphtheria, tetanus, and pertussis. MICS=Multiple Indicator Cluster Survey.

Table: **Model data sources overview**

48 International Classification of Childhood Cancer diagnoses to estimate the incidence rates of total and diagnosed cancers. We modelled the annual number of cases using our estimates of total incidence (appendix p 42). For each cancer case, we then simulated the probability of access to primary care, and the probability of appropriate referral and diagnosis, specified by the following equation:

$$\text{Diagnosed incidence} = \text{total incidence} \times \text{probability of access} \times \text{probability of referral and accurate diagnosis}$$

Calibration involves comparing model predictions with empirical data, allowing us to identify sets of parameter values that achieve a good fit.[23] We fit the model parameters using a Bayesian framework in which we assumed that the observed (diagnosed) incidence as reported in the registries is fixed, and that the model parameters that give rise to the registry data (ie, total incidence and health system barriers) are random variables. We then used model calibration to fit these parameters, identifying parameter sets (ie, combinations of parameters) that yielded model predictions of diagnosed incidence consistent with the observed data. We briefly describe this process here; the appendix (p 42) contains full details on the model calibration.

We calibrated the model to all reported country, age, and diagnosis-specific incidence rates, totalling 10 078 registry targets (ie, observed datapoints against which parameter sets were scored). We scored the model predictions based on the squared distance between the predicted and reported incidence, with each registry target weighted inversely proportional to the width of its confidence interval. For computational efficiency, we used a hybrid approach combining stochastic (simulated annealing)[24] and deterministic (gradient descent)[25] optimisation techniques to identify good-fitting parameter sets. We ran 10 000 independent searches and selected the 100 best-fitting parameter sets to account for uncertainty around the model parameters.

As a posterior predictive check[22] of our calibrated model, we compared our predictions of diagnosed incidence to registry-reported incidence. We calculated how often our prediction intervals (95% UI) contained the reported point estimate (ie, the coverage probability), and how often our mean predicted incidence fell within the 95% CIs of the registry data.

Using the best-fitting 100 parameter sets from the calibrated model to more fully portray uncertainty, we estimated the underlying total incidence rates for each diagnosis and age group. We used the WHO World Standard Population to age-standardise our reported estimates.[26] We then ran 1000 simulations to project total incident cancer cases from 2015 to 2030, taking into account trends in population growth and urbanisation. We assumed that the incidence rates and health system variables remained constant. In each simulation, we sampled a parameter set (from the top 100 parameter sets identified during calibration) and country-specific population projections. This approach allowed us to take into account both stochastic (first-order) and parameter (second-order) uncertainty to estimate the posterior predictive distributions of our model outcomes. In each iteration, we sampled a parameter set to account for parameter uncertainty, and within each iteration we simulated individual children with cancer to capture stochastic uncertainty. We ran 1000 simulations as a compromise between the computational demands of the model and the need to estimate stable means and explore parameter uncertainty. We used this Bayesian framework to estimate the number of diagnosed cases and the number of total underlying cancer cases from 2015 to 2030.

The GCC microsimulation model was coded in Java (version 1.8.0), and the statistical analyses were done in R (version 3.3.1).

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. All authors had full access to all the data used in the study. The corresponding author had final responsibility for the decision to submit for publication.

Results

Using our GCC model, we estimated that globally in 2015 there were 397 000 (95% UI 377 000–426 000) total incident cases of childhood cancer, whereas only 224 000 (95% UI 216 000–237 000) cases were diagnosed. Our estimates of annual diagnosed global childhood cancer cases are similar to estimates from the International Agency for Research on Cancer (IARC),[8] which estimated 200 000 cases in 2018, and the Global Burden of Disease Study 2016 (GBD 2016;[27] appendix p 144), which estimated 195 000 (95% UI 175 000–206 000) cases in 2016. Furthermore, our predictions of total annual childhood cancer cases are similar to recently reported totals from a convenience sample of several high-income countries where national data are available (appendix p 145). Our predictions aligned well with these estimates, with our prediction intervals containing the reported estimate for each country, building confidence in the GCC model predictions. Our posterior predictive checks (ie, comparing our predictions to all available IICC-3 registry data) revealed that nearly all (99·3%) of our prediction intervals overlapped with the 95% CIs of the registry data, and our prediction intervals contained the registry point estimate 87·7% of the time. Our mean predicted incidence fell within the registry 95% CIs 84·8% of the time over all diagnoses (appendix pp 45–143). Our highest mean squared error was for neuroblastoma (appendix p 43), where we often predict lower incidence than reported in the registries—our mean predicted incidence fell within the registry 95% CIs 79·9% of the time for this diagnosis (appendix pp 79–80).

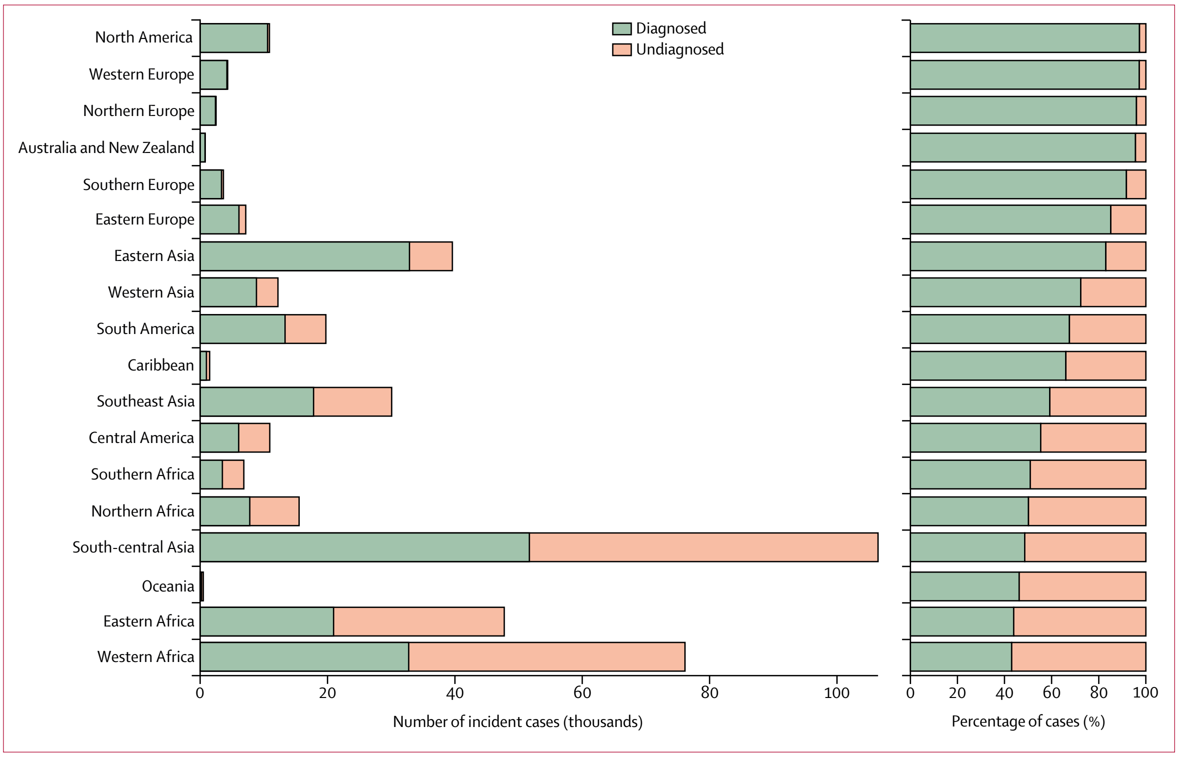

Figure 2: **Total number of incident and diagnosed cases of childhood cancer by region in 2015**

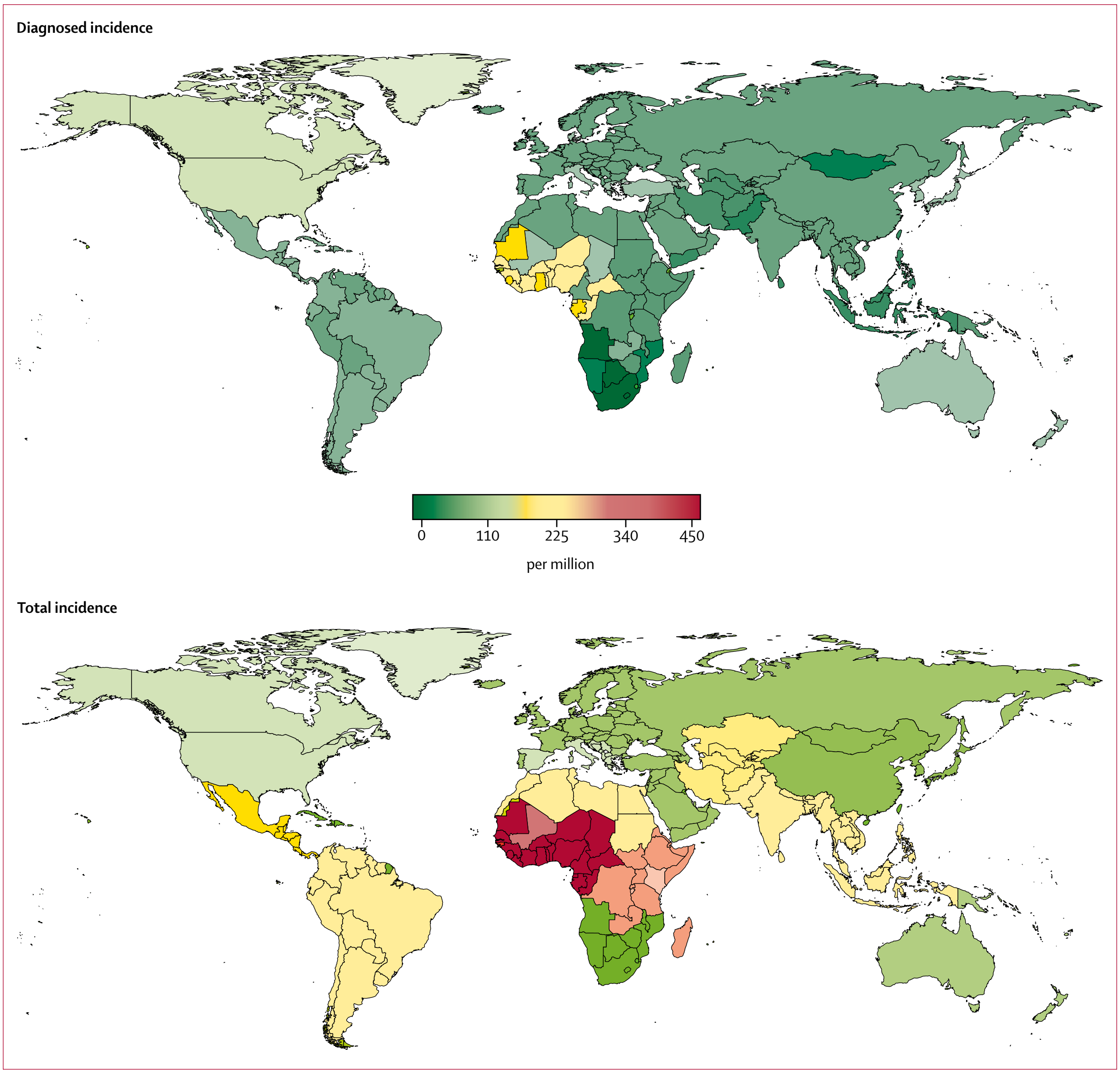

Figure 3: **Age-standardised incidence (per million) of diagnosed and total childhood cancer cases by country in 2015**

The estimated proportion of cancer cases that are underdiagnosed varies substantially by country (appendix pp 191–391). We estimated that in 2015, 43% (172 000 of 397 000) of global childhood cancer cases were not diagnosed, ranging from 3% in western Europe (120 of 4300) and North America (300 of 10 900), to 57% (43 000 of 76 000) in western Africa (figure 2). In south Asia (including southeastern Asia and south-central Asia), the overall proportion of undiagnosed cases was estimated to be 49% (67 000 of 137 000).

We estimated the total number of incident and diagnosed cases by region in 2015 (figure 2), and present the age-standardised incidence rates (per million) of diagnosed and total cases by country (figure 3). Figure 4 provides estimates of total incident cancer cases by diagnosis and region for the top 15 specified diagnoses (ie, excluding unspecified or other)—the estimated incidences for all diagnoses are presented in the appendix (pp 146–49). Estimated regional incidence rates are reported in the appendix (pp 150–61). The appendix also contains the maps of age-standardised incidence by diagnosis group (pp 162–74), and maps of estimated access and referral probabilities (p 175). We estimate that 92% (366 000 of 397 000) of total incident cases occur in low-income and middle-income countries (appendix pp 191–391).

We found that acute lymphoblastic leukaemia is the most common cancer in most regions of the world, with the notable exception of sub-Saharan Africa (including eastern, western, and southern Africa), where acute lymphoblastic leukaemia incidence is substantially lower than in other global regions

	Global	Eastern Africa	Southern Africa	Western Africa	Northern Africa	Eastern Asia	South-central Asia	Southeast Asia	Western Asia	Eastern Europe	Northern Europe	Southern Europe	Western Europe	Northern America	Carribean	Central America	South America	Oceania	Australia and New Zealand
Total incident cancer cases	396 652 (377 361–425 724)	47 753 (40 051–58 883)	6861 (5824–8011)	76 132 (60 929–96 599)	15 542 (12 700–17 545)	39 606 (33 743–48 419)	106 447 (95 078–116 965)	30 079 (25 020–34 547)	12 221 (10 947–13 706)	7165 (6351–8036)	2504 (2303–2715)	3656 (3370–3908)	4327 (4049–4798)	10 882 (10 404–11 712)	1495 (1258–1737)	10 938 (8901–12 855)	19 743 (17 209–22 112)	499 (388–620)	802 (707–930)
I(a) Acute lymphoblastic leukaemia	74 511 (66 820–83 192)	3526 (2091–5469)	823 (556–1171)	1699 (884–3307)	2173 (1268–3016)	10 173 (6421–14 711)	28 538 (23 000–34 534)	7320 (4326–12 152)	3145 (2319–4036)	1903 (1318–2520)	690 (555–827)	918 (758–1062)	1082 (913–1280)	2677 (2470–2951)	316 (157–525)	3696 (2197–4934)	5468 (4158–6718)	119 (36–201)	245 (190–318)
II(b) Non-Hodgkin lymphoma	21 661 (16 522–29 681)	3183 (1867–5814)	446 (220–751)	5888 (2465–13 459)	851 (415–1303)	1429 (542–2275)	5403 (2836–7227)	1188 (598–1859)	627 (422–852)	279 (141–416)	92 (54–126)	144 (101–184)	163 (96–231)	441 (281–555)	102 (42–196)	509 (213–761)	872 (572–1245)	14 (2–31)	30 (11–52)
VI(a) Nephroblastoma	20 978 (17 309–25 375)	3194 (2214–5220)	783 (504–1262)	6198 (2858–11 362)	974 (598–1307)	1187 (408–1942)	4164 (2704–5165)	1081 (656–1627)	542 (326–822)	463 (276–650)	126 (67–176)	176 (126–223)	238 (175–300)	550 (373–665)	82 (43–136)	340 (133–542)	822 (532–1165)	17 (3–34)	41 (14–68)
II(c) Burkitt lymphoma	19 550 (12 178–43 761)	4091 (1680–7214)	200 (120–308)	11 803 (4747–33 914)	403 (198–656)	440 (179–661)	987 (626–1340)	165 (96–273)	326 (210–459)	118 (72–186)	32 (13–51)	99 (59–135)	105 (61–144)	170 (94–232)	16 (6–29)	158 (86–233)	415 (249–595)	9 (1–19)	13 (4–24)
V Retinoblastoma	19 416 (15 025–28 512)	2310 (1540–3752)	515 (251–907)	7824 (3725–15 926)	460 (298–642)	1345 (378–2221)	3504 (2476–4785)	1393 (878–2003)	339 (216–463)	176 (105–286)	60 (32–87)	79 (51–107)	99 (65–136)	269 (195–349)	42 (20–80)	302 (144–434)	657 (461–906)	16 (2–32)	26 (11–49)
I(b) Acute myeloid leukaemia	16 905 (13 720–19 487)	673 (280–1296)	388 (259–610)	770 (395–1543)	699 (472–929)	2525 (1282–3559)	5437 (4054–6624)	2567 (1388–3899)	682 (466–907)	318 (182–464)	120 (65–162)	158 (110–203)	198 (128–253)	499 (332–610)	68 (35–119)	585 (286–833)	1128 (696–1503)	39 (7–70)	49 (27–78)
II(a) Hodgkin lymphoma	16 606 (13 834–19 616)	1901 (1236–2817)	312 (186–496)	2457 (1042–4307)	1117 (707–1488)	352 (39–605)	6407 (4074–8361)	404 (224–712)	880 (577–1224)	364 (202–505)	93 (47–129)	191 (133–249)	202 (124–269)	394 (223–505)	73 (41–118)	508 (263–730)	911 (561–1271)	13 (4–25)	26 (8–43)
IV(a) (Ganglo) neuroblastoma	14 288 (11 814–16 240)	632 (353–1233)	275 (159–567)	1247 (617–1963)	1169 (754–1564)	3007 (1545–4419)	3077 (1840–4119)	1025 (579–1558)	772 (498–1049)	150 (78–208)	545 (337–739)	261 (193–322)	299 (198–385)	743 (605–884)	93 (51–153)	239 (107–371)	675 (385–935)	24 (9–47)	56 (32–86)
III(b) Astrocytoma	14 053 (12 198–16 249)	536 (323–868)	197 (111–319)	634 (245–1395)	656 (383–953)	1620 (541–2861)	4580 (3623–5533)	954 (573–1478)	583 (337–853)	531 (277–722)	266 (165–350)	309 (221–396)	383 (239–496)	1093 (810–1287)	98 (53–163)	519 (248–784)	1032 (636–1475)	20 (5–39)	43 (16–70)
III(c) CNS embryonal	11 065 (9459–12 583)	493 (263–1034)	179 (110–284)	576 (188–1156)	571 (313–793)	1602 (697–2518)	3566 (2115–4342)	896 (558–1426)	546 (360–792)	360 (215–492)	118 (68–164)	164 (114–210)	200 (133–259)	510 (398–616)	57 (32–92)	376 (190–576)	782 (516–1069)	31 (9–57)	36 (10–64)
(IC)a Rhabdomyosarcoma	10 919 (9114–14 008)	1858 (1023–3346)	342 (193–600)	1951 (975–3929)	468 (300–652)	950 (315–1556)	2507 (1493–3082)	725 (467–1098)	377 (273–504)	210 (114–305)	74 (41–121)	116 (76–151)	145 (89–192)	323 (221–448)	37 (17–60)	238 (108–351)	560 (305–758)	11 (2–23)	28 (12–47)
(VIII)a Osteosarcoma	9572 (8264–11 063)	1165 (694–2403)	189 (95–319)	816 (436–1458)	441 (267–601)	1162 (451–2043)	2910 (2247–3576)	1074 (598–1984)	277 (175–383)	134 (80–202)	44 (21–67)	79 (46–109)	104 (66–140)	248 (142–315)	44 (22–75)	271 (89–457)	582 (357–799)	14 (3–37)	17 (4–38)
IX(c) Kaposi sarcoma	8882 (4865–16 551)	6081 (2794–13 236)	479 (173–1136)	1220 (507–3186)	231 (36–632)	103 (10–365)	183 (19–627)	66 (13–212)	26 (7–77)	5 (0–17)	5 (0–22)	3 (0–8)	3 (0–12)	2 (0–5)	26 (2–95)	151 (10–492)	256 (24–945)	20 (1–62)	21 (1–65)
X(c) Gonadal germ cell	6528 (5457–7792)	243 (130–386)	69 (28–133)	588 (265–1314)	128 (70–192)	1197 (383–1915)	2140 (1607–2676)	814 (503–1162)	175 (89–252)	94 (49–147)	30 (16–46)	51 (33–70)	58 (34–85)	145 (97–192)	26 (10–47)	324 (109–567)	426 (283–554)	9 (1–22)	12 (3–22)
VIII(c) Ewing and related	5672 (4584–6799)	404 (181–969)	44 (25–79)	553 (251–1181)	383 (250–531)	305 (127–462)	2527 (1447–3208)	209 (112–312)	286 (188–407)	121 (77–168)	41 (22–61)	76 (50–106)	97 (68–130)	162 (99–216)	16 (6–31)	108 (35–174)	306 (168–437)	16 (4–33)	17 (8–29)

0–4 | 5–9 | 10–19 | 20–34 | 35–49 | ≥50

Figure 4: **Estimated total incident cancer cases and age-standardised incidence in 2015 by the top 15 specified* diagnoses and region**
Data are mean (95% uncertainty interval). Shaded cells indicate age-standardised incidence rates per million. Differences between global values and summed regional values are due to rounding. *Top 15 diagnoses by global cases after removing other or unspecified diagnoses.

(figure 4; appendix p 150). However, we found that increased incidence of other diagnoses leads to higher overall cancer incidence in much of Africa, especially in western Africa, where we estimated the age-standardised total incidence rate to be 430 (95% UI 344–546) per million person-years compared with an average of 157 (95% UI 151–161) per million person-years in Europe and North America (figure 3, appendix pp 150–61). We found that 75% of this difference in total incidence is due to a higher incidence of lymphomas, retinoblastoma, and renal tumours in western Africa, with Burkitt lymphoma alone comprising 25% of the difference. We also found that "other" and "unspecified" cancers comprise 11% (8500 of 76 000) of cases in west Africa compared with less than 1% (200 of 29 000) in Europe and North America (figure 4, appendix pp 150–61).

Taking into account population projections, we estimated that there will be 6·7 million (95% UI 6·3–7·2) cases of childhood cancer worldwide from 2015 to 2030. At current levels of health system performance (access and referral), we estimated that 2·9 million cases (95% UI 2·7–3·3) or 44% of all childhood cancers will not be diagnosed during this period (figure 5).

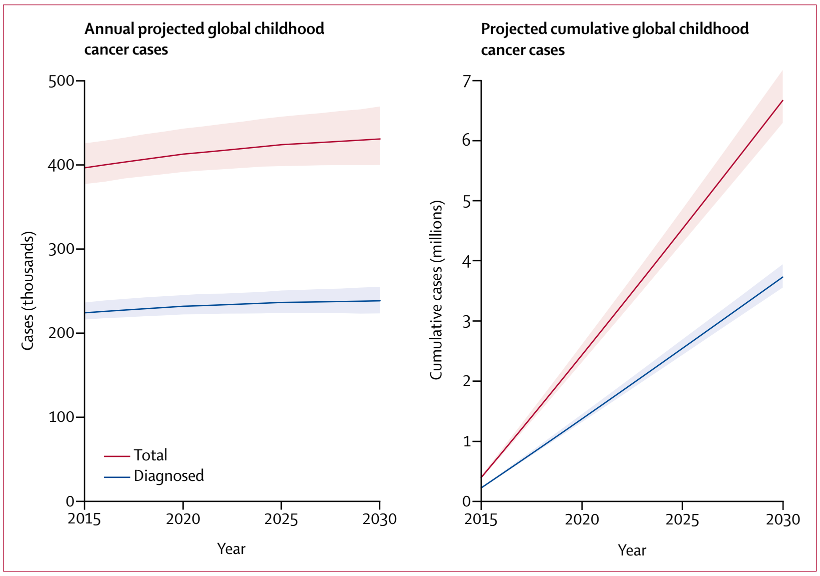

Figure 5: **Modelled projections of incident global childhood cancer cases between 2015 and 2030**
Shaded areas are 95% uncertainty intervals.

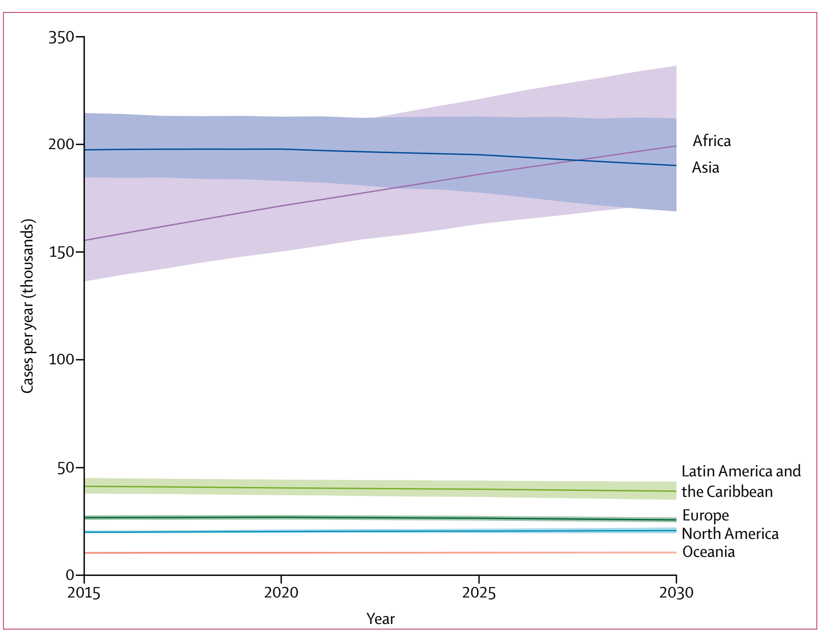

Figure 6: **Modelled projections of incident childhood cancer cases by continent between 2015 and 2030**
Shaded areas are 95% uncertainty intervals.

We also found that because of demographic trends, the number of childhood cancer cases is declining or stable in most regions of the world (figure 6). Africa is, however, a notable exception, which is projected to have substantial population growth, with the number of children aged 0–14 years increasing from 485 million in 2015 to 625 million in 2030.[14] We found that population growth in Africa will drive an increase in the number of global lymphoma cases (appendix p 176). The appendix presents projections for each country (pp 191–391).

Discussion

Using a simulation model of childhood cancer incidence in 200 countries and territories, we found that the annual global incidence of childhood cancer is about 400 000 cases after adjusting for underdiagnosis, compared with about 200 000 cases currently reported. We estimated that more than 90% of childhood cancers occur in low-income and middle-income countries—a higher proportion than previously thought. Health system barriers to access and referral result in substantial underdiagnosis of childhood cancer in many countries, with nearly one-in-two cases of global childhood cancer not diagnosed and treated.

Although our model-based estimates should be interpreted in light of data limitations and modelling assumptions, we found that our model has a high level of accuracy compared with available data; our predictions of diagnosed incidence rates are consistent with country-specific registry data and reflect geographical variation in cancer incidence and heterogeneity in health systems across and within countries, and our estimates of global diagnosed cases are similar to estimates by IARC and GBD 2016.[8,27] Furthermore, our model accurately predicted the total number of childhood cancer cases compared with national data reported for several high-income countries, and is consistent with recent findings,[28] suggesting that the increase in childhood cancer incidence observed in a subset of European cancer registries might partly reflect improvements in the diagnosis and registration of paediatric cancers over time.

In some cases, ascertainment bias in the registry data might have caused our prediction intervals to not contain the registry-reported point estimate. For example, our highest prediction error is for infant neuroblastoma incidence; we typically predict lower incidence than reported in the registries. However, because neuroblastoma has a high rate of spontaneous regression in infants,[29] countries with advanced imaging and diagnostic capabilities are able to identify a higher proportion of asymptomatic or mild symptomatic patients in this age group, suggesting that our model has good face validity in estimating clinically actionable cases of neuroblastoma.

Our findings suggest the magnitude of undiagnosed childhood cancer represents a large proportion of the total incidence, especially in south Asia and sub-Saharan Africa. Although these regions have similar proportions of undiagnosed cases, east and west Africa have registry-reported overall incidence rates as high as Europe and North America, suggesting that the underlying incidence of childhood cancer is even higher in these regions when we take into account the effect of health system barriers on diagnosis. We found that this higher overall incidence is largely due to different patterns of cancer incidence by diagnosis, with lymphomas in particular driving higher incidence in these regions.

Although the focus for global childhood cancer is typically on improving oncology care through efforts such as twinning low-income and middle-income countries with centres in high-income countries,[30] improving treatment at a small number of individual facilities, we found that interventions aimed at health system strengthening (at every step of the care cascade) will also be needed to reduce the number of undiagnosed children with cancer. Developing reliable cancer registries and health information systems will be key to monitoring progress towards the goal of identifying all cases in a population. The large magnitude of undiagnosed cancer cases presents a challenge to many countries as they increase access to childhood cancer treatment as part of universal health coverage expansion, prompted by SDG target 3.8, to "achieve universal health coverage, including financial risk protection, access to quality essential health-care services and access to safe, effective, quality and affordable essential medicines and vaccines for all".[6]

Our model-based estimates of the total incidence of childhood cancer will hopefully be able to help guide health system planning and inform new policies to improve management of childhood cancers. For example, although improving access to primary and specialty care can contribute to further reductions in child mortality (for all children, both with and without cancer), adequate cancer treatment capacity should also be planned to address the larger number of identified cancer cases. In a follow-up work, we plan to estimate the effect of improving probabilities of health system access and referral for childhood cancer, among other strategies. In this current analysis, we therefore kept these probabilities constant when projecting cancer cases between 2015 and 2030 as a baseline analysis, and to highlight the need for continued investment in health system strengthening.

Indeed, examining the underdiagnosis of childhood cancer can help shed light on health system performance. Because it is reasonable to consider the incidence of childhood cancer as a random event, the gap between total and diagnosed cases (as indicated by the coverage of paediatric cancer registry data) can in turn serve as a tracer or indicator of access and referral within a given health system. Our approach also highlights the potential importance of using structural models to estimate the effect of health system barriers on cancer diagnosis. By explicitly taking into account the population structure (age structure and urban or rural location) of children in all countries, as well as differential barriers to diagnosis (by urban or rural location within countries), we were able to fit our model to data from countries where cancer registration is more well established and then make predictions for countries without registries where the population structure and health system barriers might be different. This approach could be extended to other disease areas if data are available; the collection of IICC-3 cancer registry data over long periods of time served as the foundation for our modelling approach and highlights the importance of data collection for other diseases.

Although our modelling approach allows us to synthesise data from multiple sources in a way that is consistent with empirical data for health system barriers and reported cancer incidence, we recognise that there are limitations in the assumptions needed to develop the model. We used hierarchical models to more flexibly incorporate many assumptions, and we accounted for parameter uncertainty in all steps of developing the model. As a result, sensitivity analyses are already included in our uncertainty intervals. For example, the effect of uncertainty around future trends in population growth can be seen in the widening of our uncertainty intervals around incident cases at later timepoints. Although our results therefore incorporated various sources of uncertainty, some limitations remain.

First, although our proxy indicators for access and referral have good face validity for general health system functioning, these indicators might not be representative of childhood cancer specifically. For example, we used diarrhoeal treatment as one of our proxies for referral, which might depend on how serious an issue diarrhoea is in a given country. This variability is one reason why we selected multiple indicators as proxies for each health system barrier. Moreover, rather than directly using these estimates in the model, these proxy indicators are used to provide some sense of health system engagement by informing prior probability distributions that were sometimes substantially revised during calibration in which we aligned our model predictions with childhood cancer-specific registry data. In the interest of parsimony, we also used the same probabilities of access and referral (stratified by urban *vs* rural location) for all cancer diagnoses within a country, which might mask variation in the salience of various diagnoses that can affect these probabilities. Second, we assumed that all diagnosed cases are accurately recorded in cancer registries. In practice, however, some cases might be diagnosed but not recorded, or might be incorrectly classified because of deficient pathology services, which require expertise and access to immunohistochemistry. Third, although we used hierarchical models to incorporate all available data, our results might be affected by small sample sizes in some regions. For example, there were only two countries in west Africa (Mali and Cameroon) with available registry data, so our predictions for this region might be influenced by the extent to which these countries are representative of the region as a whole. Although our estimates of the proportion of undiagnosed cancer cases in Africa are similar to those in other regions, our resulting estimates of underlying incidence rates for some diagnoses in Africa might be insufficiently regularised given the small number of available registries, and thus could be overestimates. Registry data from additional countries would help to better account for potential heterogeneity and control for outliers in cancer incidence within regions. Although in general our hierarchical approach helps to guard against overfitting to outliers by regularising estimates from individual countries, this bias–variance trade-off results in posterior shrinkage in which our predictions have smaller between-country variance than the reported estimates. However, as new country-specific data become available, our model can be refined to provide updated estimates. Fourth, the registry data we used to calibrate the model pooled incidence rates over all reported years. For example, in the USA, we fit to incidence rates estimated for the period 1998–2012. Although pooling helps provide more stable estimates, it could mask trends in incidence rates within the reported period, such as those due to changing demographic composition of racial or ethnic subgroups with different cancer risks. Lastly, we did not account for competing causes of death in this analysis, which would reduce the number of prevalent undiagnosed cancer cases. However, competing mortality should not affect the number of undiagnosed incident cases (at least relative to diagnosed cases), which is the focus of this analysis. Estimating the prevalence of childhood cancer would require simulating incidence and mortality (from all causes) for a specified time period, which is the focus of a follow-up analysis.

Notwithstanding these limitations, using a model-based approach, we provide, to our knowledge, the first global estimate of the extent of underdiagnosis of childhood cancer and estimate the total underlying incidence of each International Classification of Childhood Cancer diagnosis. We found that health system barriers result in substantial underdiagnosis of childhood cancer cases in many countries. The true incidence of global childhood cancer is likely to be about 400 000 cases per year—substantially higher than the number currently reported—indicating that nearly one-in-two cases of childhood cancers is not diagnosed and treated.

In conclusion, our findings highlight the need for continued investment in health systems to address the underdiagnosis and large hidden incidence of childhood cancer, and the importance of expanding cancer registration to track progress towards the goal of universal access and treatment. As health systems expand access to childhood cancer care as part of universal health coverage, these estimates could help to guide the development of policies to ensure effective access to care for all children with cancer.

Contributors

All authors designed the study. ZJW acquired the data for the study and did the analyses. All authors interpreted the results and contributed to the writing of the report.

Declaration of interests

We declare no competing interests.

Acknowledgments

This study was funded by the Boston Children's Hospital, Dana-Farber Cancer Institute, Harvard T H Chan School of Public Health, Harvard Medical School, National Cancer Institute, SickKids, St Jude Children's Research Hospital, and the Union for International Cancer Control.

References

1 Global Burden of Disease Cancer Collaboration, Fitzmaurice C, Allen C, et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: a systematic analysis for the Global Burden of Disease Study. *JAMA Oncol* 2017; **3:** 524–48.

2 Bonaventure A, Harewood R, Stiller CA, et al. Worldwide comparison of survival from childhood leukaemia for 1995–2009, by subtype, age, and sex (CONCORD-2): a population-based study of individual data for 89 828 children from 198 registries in 53 countries. *Lancet Haematol* 2017; **4:** e202–17.

3 Magrath I, Steliarova-Foucher E, Epelman S, et al. Paediatric cancer in low-income and middle-income countries. *Lancet Oncol* 2013; **14:** e104–16.

4 Rodriguez-Galindo C, Friedrich P, Alcasabas P, et al. Toward the cure of all children with cancer through collaborative efforts: pediatric oncology as a global challenge. *J Clin Oncol* 2015; **33:** 3065–73.

5 Farmer P, Frenk J, Knaul F, et al. Expansion of cancer care and control in countries of low and middle income. *Lancet* 2010; **376:** 1186–93.

6 UN General Assembly. Transforming our world: the 2030 agenda for sustainable development. 2015. http://www.refworld.org/docid/57b6e3e44.html (accessed Aug 6, 2018).

7 Steliarova-Foucher E, Colombet M, Ries LAG, et al. International incidence of childhood cancer, 2001–10: a population-based registry study. *Lancet Oncol* 2017; **18:** 719–31.

8 Ferlay J, Colombet M, Soerjomataram I, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer* 2019; **144:** 1941–53.

9 Bray F, Colombet M, Mery L, et al. Cancer incidence in five continents volume XI. Lyon: International Agency for Research on Cancer, 2017.

10 Valsecchi MG, Steliarova-Foucher E. Cancer registration in developing countries: luxury or necessity? *Lancet Oncol* 2008; **9:** 159–67.

11 Zhang J, Walsh MF, Wu G, et al. Germline mutations in predisposition genes in pediatric cancer. *N Engl J Med* 2015; **373:** 2336–46.

12 Gröbner SN, Worst BC, Weischenfeldt J, et al. The landscape of genomic alterations across childhood cancers. *Nature* 2018; **555:** 321–27.

13 World Bank. World Bank country and lending groups. https://datahelpdesk.worldbank.org/knowledgebase/articles/906519 (accessed May 17, 2018).

14 UN, Department of Economic and Social Affairs, Population Division. World Population Prospects. Volume I: comprehensive tables (2015 revision). https://population.un.org/wpp/Publications/Files/WPP2015_Volume-I_Comprehensive-Tables.pdf (accessed Feb 20, 2019).

15 UN, Department of Economic and Social Affairs, Population Division. World Population Prospects. Volume II: demographic profiles (2015 revision). https://population.un.org/wpp/Publications/Files/WPP2015_Volume-II-Demographic-Profiles.pdf (accessed Feb 20, 2019).

16 UN, Department of Economic and Social Affairs, Population Division. World Urbanization Prospects: the 2014 revision, CD-ROM edn.

17 Steliarova-Foucher E, Colombet M, Ries LAG, et al. International incidence of childhood cancer, volume 3. 2017. Lyon, France: International Agency for Research on Cancer. http://iicc.iarc.fr/results/ (accessed May 17, 2018).

18 WHO. Global Health Observatory data repository. Geneva: World Health Organization. http://apps.who.int/gho/data/node.home (accessed May 17, 2018).

19 Steliarova-Foucher E, Stiller CA, Lacour B, Kaatsch P. International classification of childhood cancer, third edition. *Cancer* 2005; **103:** 1457–67.

20 Hogan DR, Stevens GA, Hosseinpoor AR, Boerma T. Monitoring universal health coverage within the Sustainable Development Goals: development and baseline data for an index of essential health services. *Lancet Glob Health* 2018; **6:** e152–68.

21 WHO. Global reference list of 100 core health indicators, 2015. Geneva: World Health Organization, 2015. http://www.who.int/healthinfo/indicators/2015/en/ (accessed Aug 21, 2018).

22 Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis, 3rd edn. Boca Raton, FL: CRC Press, 2014.

23 Vanni T, Karnon J, Madan J, et al. Calibrating models in economic evaluation: a seven step approach. *Pharmacoeconomics* 2011; **29:** 35–49.

24 Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by simulated annealing. *Science* 1983; **220:** 671–80.

25 Forsyth D. Probability and statistics for computer science. Cham, Switzerland: Springer International Publishing, 2018.

26 WHO. WHO world standard population. Geneva: World Health Organization. http://apps.who.int/healthinfo/statistics/mortality/whodpms/definitions/pop.htm (accessed May 17, 2018).

27 Institute for Health Metrics and Evaluation. Global Burden of Disease Study 2016 results. 2017. http://ghdx.healthdata.org/gbd-results-tool (accessed May 17, 2018).

28 Steliarova-Foucher E, Fidler MM, Colombet M, et al. Changing geographical patterns and trends in cancer incidence in children and adolescents in Europe, 1991–2010 (Automated Childhood Cancer Information System): a population-based study. *Lancet Oncol* 2018; **19:** 1159–69.

29 Brodeur GM. Spontaneous regression of neuroblastoma. *Cell Tissue Res* 2018; **372:** 277–86.

30 Sullivan R, Kowalczyk JR, Agarwal B, et al. New policies to address the global burden of childhood cancers. *Lancet Oncol* 2013; **14:** e125–35.