Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The origins of childhood cancer

qiaoxuan wei
3586791320qq.com
The Fourth People's Hospital of Nanning https://orcid.org/0009-0003-7824-0311
yuting ma
The Fourth People's Hospital of Nanning https://orcid.org/0009-0006-8968-0111
guijuan zhu
The Fourth People's Hospital of Nanning https://orcid.org/0009-0000-2302-9912

Research Article

Keywords: childhood cancer, developmental sequence, gate failure, latent-variable model, age-specific incidence, silent clone, autopsy data

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11626154/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Background

Treatment success in childhood cancer has reached a plateau. Aetiological research remains at the level of molecular inventories. The first fact is that the window for mutation accumulation is measured in months. The second fact is that no recognisable multi-step precancerous sequence exists in the general paediatric population. The third fact is that tumour histology follows developmental programmes. Existing models answer where the field is, what it looks like and which lineage it comes from. These models do not provide a measurable state. That state should distinguish cells that will complete transformation from cells that will not.

Methods

We built a five-state latent-variable developmental model. The P state is persistent fetal developmental-sequence deviation. The G state is failure of growth-phase epigenetic correction and gate failure. The C state is a gradient-cell clone. The T state is a clinical tumour. The R state is a regressed or silent clone. The model was calibrated to age-specific incidence for five developmental-origin subtypes in SEER 2008–2012. The five subtypes were neuroblastoma, retinoblastoma, Wilms tumour/renal tumour, hepatoblastoma and embryonal central nervous system tumour. The model used autopsy silent-clone data to anchor P and R. The model used a single-branch mutation-accumulation model as a comparator. The model simulated the effect of lowering the gate-failure rate or raising the clone-regression rate within the developmental window.

Results

The model reproduced age-specific incidence for the five subtypes. Observed and predicted peak groups agreed in four of five subtypes. The observed peak for embryonal central nervous system tumour was < 1 year. The predicted peak for this subtype was 1–4 years. The two peaks differed by one age group. The predicted composite incidence was 107.6/79.8/16.6/5.2/2.3 per million. The observed composite incidence was 115.9/65.0/16.6/5.6/2.6. The latent-variable model had a lower AIC than the single-branch model in five of five subtypes. ΔAIC was 3.1–416.9. The clone-to-clinical ratio was 37.6–72. The model value for neuroblastoma was 0.66% clone and 0.018% clinical. Autopsy in situ lesions were 1/150–1/250. Clinical neuroblastoma was 1/7,000. The model ratio was in the same direction as the autopsy ratio. A 50% rise in kR reduced cumulative clinical incidence by 24.4%–26.7% across the five subtypes. A 50% fall in kG0 had an effect of 0%–21.1%.

Conclusions

The rate of mutation accumulation cannot produce the early peak of childhood cancer. The single-branch model cannot reproduce the peak groups of the five subtypes. The entry point of childhood cancer is persistent fetal developmental-sequence deviation, growth-phase correction failure and gate failure. The clinical tumour is the terminal readout of this path. The clone-regression rate within the window is an earlier intervention variable than the mutation rate.

1 Introduction

1.1 Treatment success and the aetiological gap

Childhood cancer is the most successfully treated group of malignancies in oncology. Five-year overall survival for ages 0–19 in the United States rose from 60% in 1975 to 90% in 2017 [1]. Australian registry data on 20,871 cases show five-year relative survival rising from 72.8% in 1983–1994 to 86.1% in 2007–2016 [2]. Cure in acute lymphoblastic leukaemia exceeds 90% [3]. Global totals in 2023 were 377,000 new cases and 144,000 deaths in ages 0–19 [4].

Treatment success has not answered the aetiological question. Relapse remains the leading cause of death. Relapse concentrates among children initially classified as standard risk [3]. Treatment intensity has reached a plateau. A randomised trial of 630 patients compared two induction regimens. The two regimens did not differ in efficacy. The stronger regimen had significantly more grade 3–4 toxicity [36]. 6-Thioguanine replaced 6-mercaptopurine. The replacement increased toxicity. The replacement produced no benefit [37]. Standard-risk acute lymphoblastic leukaemia lowered delayed-intensification intensity. The reduction did not prove non-inferiority [38]. The cost is deferred to survivors. Among 10,397 adult survivors, chronic health conditions reached 73.4% at 30 years after diagnosis [39]. Grade 3–5 conditions in these survivors reached 96.0% by age 50 [40]. Forty-year all-cause mortality among 34,230 five-year survivors reached 23.3% [41].

These three facts point to a problem that the current system ignores. The treatment target is the tumour that has already appeared. The current system does not ask why this child had the opportunity to begin. Every quantity measured by the current system takes the presence of tumour as its object. The current system has no site that reads the constraint state of a cell before it is allowed to grow.

1.2 Three anomalies that adult templates cannot accommodate

Childhood cancer aetiology has three anomalies that adult templates cannot accommodate. The first anomaly is that the window for mutation accumulation is measured in months. The initiating event occurs in utero. Transformation occurs shortly after birth [5][6]. Exposure accumulation in adult epithelial cancer is measured in decades. Childhood cancer does not have this time.

The second anomaly is that no recognisable multi-step precancerous sequence exists in the general paediatric population. Precursor lesions that have been found are confined to developmental and germline contexts. Nephrogenic rests are precursor lesions of Wilms tumour. Isolated rests progressed in 33% of an observation group. Chemotherapy reduced the progression rate of isolated rests to 3.9%. Multifocal rests progressed at 50%. Diffuse hyperplastic nephroblastomatosis progressed at 100% [7]. In situ neuroblastoma can be found incidentally at autopsy [15]. Transient abnormal myelopoiesis in Down syndrome is a neonatal pre-leukaemic state that can be detected actively [8]. Investigators found these lesions only after prior marking of a known developmental or germline context. Current criteria do not read the corresponding state in the general population.

The third anomaly is that tumours follow developmental programmes. Inactivation of the same gene is oncogenic only in a specific cell at a specific developmental stage. With RB loss, maturing human cone precursors enter the cell cycle. These cells form tumours. Immature cone precursors do not form tumours [9][10]. Childhood tumour histology corresponds precisely to a developmental stage. Adult tumour histology often lies at a clear distance from the tissue of origin.

1.3 Where existing models stand

The entry point of childhood cancer is not mutation accumulation. The somatic mutation-accumulation model produced the molecular inventory. This model showed that childhood cancers carry few somatic mutations. This model showed that childhood cancers carry a higher proportion of germline change [11]. Research on in utero origin turned the seed from conjecture into evidence [5][6]. The two-step model supplied the structure of intrauterine initiation followed by postnatal secondary change [5]. The developmental-window, cell-of-origin and epigenetic model is closest to the position taken here. This model matched tumour subgroups to specific developmental lineages [12]. All four models answered where the field is, what it looks like and which lineage it comes from. None of these four models provided a measurable state. That state should distinguish cells that will complete transformation from cells that will not.

1.4 The model and contribution of this paper

We wrote the entry point of childhood cancer as a five-state path. P is persistent fetal developmental-sequence deviation. G is failure of growth-phase epigenetic correction and gate failure. C is a gradient-cell clone. T is a clinical tumour. R is a regressed or silent clone.

This path provides three operational items. First, the entry point has a state. P is a developmental programme that has already deviated at birth. G is the correction and gate construction that growth has not completed. Second, the state has readouts. The readouts are age-specific incidence, autopsy silent-clone proportion and the clone-to-clinical ratio. Third, the state has materials. The materials are SEER registries, autopsy series, pathology archives and developmental-lineage experiments.

Our contribution is to write this path as a parameterised latent-variable model. We calibrated this model to public data. This model is a latent-variable model. This model is not a molecular-mechanism model. The model places the molecular layer downstream. The model answers three questions. The first question is why the opportunity to begin exists. The second question is why the peak falls at a developmental age. The third question is why a multi-step precancerous sequence is absent.

2 Methods

2.1 Data sources

Table 1 lists data sources and uses. Age-specific incidence came from SEER Cancer Statistics Review 1975–2012, Table 29.1. The time window of this table was 2008–2012. The grouping of this table was ICCC group. The unit of this table was per million child-years [14]. International incidence of childhood cancer (CI5) provided a second test of the age distribution for the same subtypes [13]. We used SEER for calibration. We used CI5 only for external consistency. We selected five developmental-origin subtypes. The five subtypes were neuroblastoma and ganglioneuroblastoma (ICCC IVa), retinoblastoma (V), nephroblastoma and other non-epithelial renal tumours (VIa), hepatoblastoma (VIIa) and intracranial or intraspinal embryonal tumours (IIIc).

Autopsy silent-clone data came from three groups. The first group was in situ neuroblastoma. The detection rate of this lesion in stillbirth and infant autopsy series was 1/150–1/250. Clinical neuroblastoma was 1/7,000 newborns [15]. The second group was nephrogenic rests. These rests were present in up to 1% of routine infant autopsies. Most nephrogenic rests regressed within the first year. The annual incidence of Wilms tumour was about 10 per million children [7][16]. The third group was transient abnormal myelopoiesis (TAM) in infants with Down syndrome. TAM occurred in about 10%. GATA1 mutations were detectable in up to 30%. Fewer than 25% of TAM cases progressed to myeloid leukaemia [17][8].

Developmental-lineage and gate-mechanism evidence came from published human and experimental studies. Human cone precursors with RB1 loss provided developmental-stage-specific evidence [9][10]. Cerebellar granule neuron precursors and BMP/Atoh1 provided positional-information evidence [30][31]. Medulloblastoma subgroups provided lineage evidence [32][33]. ETV6-RUNX1 provided in utero origin evidence [5][6][34]. WT1 in

Wilms tumour provided renal developmental-programme evidence [35]. The intrinsic gate had four execution paths [18][19][20][21][22]. Membrane potential provided the differentiation readout [25][26][27][28].

Table 1
Data sources and uses.

Data	Source	Use
Age-specific incidence	SEER CSR 1975–2012, Table 29.1, 2008–2012 [14]	Calibration of five subtypes and composite incidence
In situ neuroblastoma	Infant autopsy series, 1/150–1/250 [15]	Anchor P and R
Nephrogenic rests	Infant autopsy, up to 1% [7][16]	Anchor P and R
TAM	Down syndrome newborns, about 10% [17][8]	Anchor P and progression proportion
Developmental lineage	Cone precursors, cerebellar granule precursors, metanephric mesenchyme [9][10][30][35]	Set aG
Gate execution paths	OIS, apoptosis, replication stress, ferroptosis [18]–[22]	Define the biological content of G
Membrane potential	Skeletal muscle, liver cancer stem cells, fetal neural stem cells [25]–[28]	Define the differentiation readout

2.2 Five-state model

Age is denoted a and measured in years. The states are P, G, C, T and R. At birth, P(0)=p0. The initial values of the other states are zero. The transition intensity from P to G is hG(a). hG(a) is a developmental-window function:

$$hG(a) = kG0 \exp[-(a - aG)^2 / (2\sigma G^2)].$$

The transition intensity from G to C is kC. The transition intensity from C to T is kT. The transition intensity from C to R is kR. The continuous-time equations are:

$$dP/da = - hG(a)\, P,$$

$$dG/da = hG(a)\, P - kC\, G,$$

$$dC/da = kC\, G - (kT + kR)\, C,$$

$$dT/da = kT\, C,$$

$$dR/da = kR\, C.$$

The clinical incidence at age a is λ(a). λ(a) equals kT multiplied by C(a). The model was discretised in monthly steps. Transition probabilities were calculated as 1 − exp(− rate × Δt). The predicted incidence for an age group was the person-year-weighted average of λ(a) within that group. This average was multiplied by 10^6 and compared with the SEER rate per million child-years.

2.3 Parameterisation and calibration

Each subtype had five free parameters. The five parameters were p0, kG0, kC, kT and kR. aG and σG were set from published developmental windows of the cells of origin. aG and σG did not enter the fit. The neuroblastoma values were aG = 0.5 and σG = 0.40. The retinoblastoma values were aG = 0.4 and σG = 0.25. The Wilms tumour values were aG = 2.5 and σG = 1.20. The hepatoblastoma values were aG = 0.5 and σG = 0.50. The embryonal central nervous system tumour values were aG = 3.0 and σG = 2.00.

Parameter estimation used penalised least squares. The objective summed the squared difference between log predicted rate and log observed rate. The sum was divided by the measurement-error variance σ = 0.15. The objective added a heavy-tailed prior. The prior settings were as follows. The mean of log p0 was log(0.005), and the standard deviation was 1.2. The mean of log kG0 was log(5), and the standard deviation was 1.2. The mean of log kC was log(0.5), and the standard deviation was 1.2. The mean of log kT was log(0.1), and the standard deviation was 1.2. The mean of log kR was log(5), and the standard deviation was 1.2. Age groups with an observed value of zero were treated as below the detection limit. These age groups did not enter the likelihood. A zero observation did not penalise the predicted value. Optimisation used 80,000 random starts and coordinate descent. Optimisation selected the parameter set with the smallest objective.

2.4 Model comparison

The single-branch mutation-accumulation model was λM(a)=λ0 a^γ. This model had two parameters. The two parameters were λ0 and γ. This model was fitted to the same positive observations. We compared the AIC of the latent-variable model and the single-branch model. AIC equals 2k plus 2NLL. NLL equals 0.5 times Σ[log λpred − log λobs]² divided by σ², plus n log σ. k is the number of parameters. The latent-variable model had k = 5.

The single-branch model had k = 2. n is the number of positive observations. The AIC difference tested whether developmental-window parameters were necessary.

2.5 Silent clones and intervention simulation

The model output three quantities. The first quantity was the cumulative probability of entering C before age 20. The second quantity was the cumulative probability of entering T before age 20. The third quantity was the ratio of the first two quantities. Cumulative clone corresponded to silent lesions visible at autopsy. Cumulative clinical corresponded to registry incidence. The model compared the clone-to-clinical ratio with the autopsy-to-clinical ratio.

The intervention simulation changed two parameters. The first change lowered kG0 by 25% and 50%. This change represented a lower gate-failure rate within the window. The second change raised kR by 25% and 50%. This change represented a higher clone-regression rate. Each change recalculated the cumulative clinical probability before age 20. The model reported the percentage fall from baseline.

2.6 Software and reproducibility

The model, fit, result tables and figures were implemented in Node.js. The code files were etaz_cp_model.js, etaz_cp_fit2.js, etaz_cp_results.js, etaz_cp_tables.js and make_etaz_figs.js. The result files were etaz_cp_fit_results.json and etaz_cp_results.json. All input data came from public registries and published literature. This study used no individual-level data.

3 Results

3.1 The model reproduces age-specific incidence for five subtypes

Table 2 gives the fitted parameters. The range of p0 was 0.276%–0.663%. The range of kG0 was 2.50–9.53 per year. The range of kC was 0.124–1.227 per year. The range of kT was 0.055–0.133 per year. The range of kR was 4.73–6.46 per year. kR was 47–86 times larger than kT. This ratio indicates that most clones regress before they reach clinical detection.

Subtype	aG (years)	σG (years)	p0	kG0 (/year)	kC (/year)	kT (/year)	kR (/year)
Neuroblastoma	0.5	0.40	0.663%	7.74	0.415	0.133	6.21
Retinoblastoma	0.4	0.25	0.396%	4.56	1.227	0.080	5.64
Wilms tumour/renal tumour	2.5	1.20	0.541%	5.43	0.428	0.107	4.99
Hepatoblastoma	0.5	0.50	0.276%	2.50	1.139	0.055	4.73
Embryonal central nervous system tumour	3.0	2.00	0.608%	9.53	0.124	0.124	6.46

Table 2. Fitted parameters of the five-state model. p0 is the proportion of the P state at birth. kG0 is the P→G intensity within the developmental window. kC is the G→C intensity. kT is the C→T intensity. kR is the C→R intensity. aG and σG were set from developmental-lineage literature and did not enter the fit.

Table 3 gives observed and predicted values. The observed values for neuroblastoma were 49.2/20.5/3.8/1.1/0. The predicted values for neuroblastoma were 41.1/27.4/4.5/0.6/0.1. The observed values for retinoblastoma were 29.2/8.7/0/0/0. The predicted values for retinoblastoma were 30.0/8.5/0.1/0/0. The observed values for Wilms tumour were 14.3/18.8/5.4/0.6/0. The predicted values for Wilms tumour were 13.5/25.1/4.3/0.5/0.1. The observed values for hepatoblastoma were 11.0/6.5/0/0/0. The predicted values for hepatoblastoma were 12.9/5.6/0.1/0/0. The observed values for embryonal central nervous system tumour were 12.2/10.5/7.4/3.9/2.6. The predicted values for embryonal central nervous system tumour were 10.1/13.3/7.7/4.1/2.2.

Subtype	Observed <1	Observed 1–4	Observed 5–9	Observed 10–14	Observed 15–19	Predicted <1	Predicted 1–4	Predicted 5–9	Predicted 10–14	Predicted 15–19
Neuroblastoma	49.2	20.5	3.8	1.1	0.0	41.1	27.4	4.5	0.6	0.1
Retinoblastoma	29.2	8.7	0.0	0.0	0.0	30.0	8.5	0.1	0.0	0.0
Wilms tumour/renal tumour	14.3	18.8	5.4	0.6	0.0	13.5	25.1	4.3	0.5	0.1
Hepatoblastoma	11.0	6.5	0.0	0.0	0.0	12.9	5.6	0.1	0.0	0.0
Embryonal central nervous system tumour	12.2	10.5	7.4	3.9	2.6	10.1	13.3	7.7	4.1	2.2
Composite	115.9	65.0	16.6	5.6	2.6	107.6	79.8	16.6	5.2	2.3

Table 3. Observed and predicted age-specific incidence for the five subtypes (per million child-years). Observed values came from SEER 2008–2012 [14]. Predicted values came from the five-state model.

3.2 Developmental windows predict peak groups

Table 4 and Figure 3 give observed and predicted peaks. The observed peak and the predicted peak agreed for neuroblastoma. The observed peak and the predicted peak agreed for retinoblastoma. The observed peak and the predicted peak agreed for Wilms tumour. The observed peak and the predicted peak agreed for hepatoblastoma. The observed peak for embryonal central nervous system tumour was <1 year. The predicted peak for embryonal central nervous system tumour was 1–4 years. The observed rates in the two age groups were 12.2 and 10.5 per million. The observed rates in the two age groups differed by 1.7 per million. Peak groups agreed in four of five subtypes.

Subtype	Observed peak	Predicted peak	Agreement
Neuroblastoma	<1	<1	Yes
Retinoblastoma	<1	<1	Yes
Wilms tumour/renal tumour	1–4	1–4	Yes
Hepatoblastoma	<1	<1	Yes
Embryonal central nervous system tumour	<1	1–4	No

Table 4. Observed and predicted peak age groups.

Sensitivity analysis moved aG up and down by 0.5 years. Sensitivity analysis moved σG up and down by 25%. Neuroblastoma and retinoblastoma moved to 1–4 years when aG was raised by 0.5 years. The other subtypes kept their peak groups. Peak order was insensitive to σG.

3.3 The silent-clone reservoir and autopsy data

Table 5 and Figure 4 give cumulative clone, cumulative clinical and the clone-to-clinical ratio. The range of cumulative clone was 0.256%–0.662%. The range of cumulative clinical was 0.004%–0.018%. The range of the clone-to-clinical ratio was 37.6–72. The model values for neuroblastoma were 0.662% clone and 0.018% clinical. The clone-to-clinical ratio for neuroblastoma was 37.6. Autopsy in situ lesions were 1/150–1/250. This proportion was 0.4%–0.67%. Clinical neuroblastoma was 1/7,000. This proportion was 0.014%. The autopsy ratio was about 29–48. The model values for Wilms tumour were 0.541% clone and 0.014% clinical. Nephrogenic rests were present in up to 1% of infant autopsies. The annual incidence of Wilms tumour was about 10 per million [7][16].

Subtype	p0	Cumulative clone	Cumulative clinical	Clone/clinical ratio
Neuroblastoma	0.663%	0.662%	0.018%	37.6
Retinoblastoma	0.396%	0.370%	0.006%	57.7
Wilms tumour/renal tumour	0.541%	0.541%	0.014%	39.2
Hepatoblastoma	0.276%	0.256%	0.004%	72.0
Embryonal central nervous system tumour	0.608%	0.556%	0.013%	41.7

Table 5. The silent-clone reservoir: cumulative clone and cumulative clinical probability before age 20, and the clone-to-clinical ratio.

3.4 The single-branch mutation model fails

Table 6 and Figure 5 give the AIC comparison. The single-branch model predicted monotonic rise of incidence with age in all five subtypes. The single-branch model could not produce an early peak. The AIC of the latent-variable model was 20.8, 2.5, 2.4, 4.4 and −3.6. The AIC of the single-branch model was 437.8, 39.4, 361.1, 7.5 and 95.0. ΔAIC was 416.9, 36.9, 358.8, 3.1 and 98.5. The latent-variable model won in all five subtypes.

Subtype	Latent-variable AIC	Single-branch AIC	ΔAIC
Neuroblastoma	20.8	437.8	416.9
Retinoblastoma	2.5	39.4	36.9
Wilms tumour/renal tumour	2.4	361.1	358.8
Hepatoblastoma	4.4	7.5	3.1
Embryonal central nervous system tumour	−3.6	95.0	98.5

Table 6. AIC of the latent-variable model and the single-branch mutation model. ΔAIC is single-branch minus latent-variable. A positive value indicates a better latent-variable model.

3.5 Window intervention simulation

Table 7 and Figure 6 give the intervention simulation. A 25% rise in kR reduced cumulative clinical incidence by 14.7%–16.1%. A 50% rise in kR reduced this incidence by 24.4%–26.7%. A 25% fall in kG0 had an effect of 0%–7.2%. A 50% fall in kG0 had an effect of 0%–21.1%. Neuroblastoma and Wilms tumour were insensitive to a fall in kG0. Clone formation in these two subtypes was close to saturation. Retinoblastoma, hepatoblastoma and embryonal central nervous system tumour were more sensitive to a fall in kG0.

Subtype	kG0 −25%	kG0 −50%	kR +25%	kR +50%
Neuroblastoma	0.5%	3.0%	14.8%	24.7%
Retinoblastoma	6.9%	20.6%	15.4%	25.5%
Wilms tumour/renal tumour	0.0%	0.0%	15.7%	26.2%
Hepatoblastoma	7.2%	21.1%	16.1%	26.7%
Embryonal central nervous system tumour	0.1%	0.3%	14.7%	24.4%

Table 7. Window intervention simulation: relative fall in cumulative clinical incidence.

3.6 Composite incidence and model boundaries

Composite incidence was close to observation in four age groups: <1, 5–9, 10–14 and 15–19 years. The predicted values in these four age groups were 107.6, 16.6, 5.2 and 2.3. The observed values in these four age groups were 115.9, 16.6, 5.6 and 2.6. The predicted value in the 1–4 year group was 79.8. The observed value in the 1–4 year group was 65.0. The difference came from Wilms tumour and embryonal central nervous system tumour. The predicted value for Wilms tumour was 25.1, and the observed value was 18.8. The predicted value for embryonal central nervous system tumour was 13.3, and the observed value was 10.5. The predicted curves for these two subtypes were broad at 1–4 years. The broad peak was the main residual of the model.

Peak groups agreed in four of five subtypes. The observed peak for embryonal central nervous system tumour was <1 year. The predicted peak for embryonal central nervous system tumour was 1–4 years. The observed rates in the two age groups differed by 1.7 per million. The continuous peak of the model lay at 1–2 years. This difference was one age group. This difference did not change peak order. Peak order was neuroblastoma/retinoblastoma < hepatoblastoma < Wilms tumour < embryonal central nervous system tumour.

3.7 Data anchors

Table 8 places each model component on a verifiable data anchor. The anchor for P was autopsy in situ lesions and nephrogenic rests. The anchor for G was developmental windows and gate execution paths. The anchor for C was gradient cells and the clone-to-clinical ratio. The anchor for T was SEER age-specific incidence. The anchor for R was autopsy silent clones and spontaneous regression of TAM.

Component	Data anchor	Readout	Source
P persistent fetal-sequence deviation	In situ neuroblastoma 1/150–1/250; nephrogenic rests ≤1%	Silent lesions present at birth	[15][7][16]
G correction failure/gate failure	Cerebellar granule precursor BMP/Atoh1; RB1 cone precursor; WT1 metanephric mesenchyme	Developmental window and lineage specificity	[9][10][30] [35]
C gradient-cell clone	Neuroblastoma SCP-like subclone; GATA1 clone in TAM	Clone present, most do not transform	[8][17]
T clinical tumour	SEER 2008–2012 age-specific incidence	Per million child-years	[14]
R regressed/silent clone	In situ neuroblastoma; spontaneous TAM regression; nephrogenic rest regression	Clone-to-clinical ratio 37.6–72	[15][17][7]
Gate execution paths	OIS, apoptosis, replication stress, ferroptosis	Four independent termination arms	[18]–[22]
Differentiation readout	Skeletal muscle <1 mV across ages 1–99; liver cancer stem cells depolarised; fetal neural stem cells −30 mV	Membrane-potential fall	[25]–[28]

Table 8. Model components and data anchors.

4 Discussion

4.1 Main findings

The five-state latent-variable model reproduced age-specific incidence for five developmental-origin subtypes. The model also reproduced the peak groups of the five subtypes. Peak groups agreed in four of five subtypes. Embryonal central nervous system tumour differed by one age group. Composite incidence was close to observation in four age groups. The 1–4 year group was overestimated because of broad peaks in Wilms tumour and embryonal central nervous system tumour. The latent-variable model had a lower AIC than the single-branch mutation model in five of five subtypes. ΔAIC was 3.1–416.9. The clone-to-clinical ratio of the model was 37.6–72. The autopsy-to-clinical ratio for neuroblastoma was 29–48. The nephrogenic-rest/Wilms-tumour ratio was in the same direction.

These three results point to one conclusion. Age-specific incidence of childhood cancer is a readout of developmental windows. Age-specific incidence of childhood cancer is not a readout of mutation-accumulation rate. The single-branch model could not produce an early peak. The single-branch model could not produce subtype order. The latent-variable model used one developmental-window parameter per subtype to produce both peak and order.

4.2 Relation to existing models

The somatic mutation-accumulation model answers which tumour, where it grows and how it progresses [11]. Research on in utero origin answers where the seed comes from [5][6]. The two-step model answers what the second hit is [5]. The developmental-window and cell-of-origin model answers where the field is [12]. This paper answers why the opportunity to begin exists, why the peak falls at a developmental age, and why a multi-step precancerous sequence is absent.

Clonal haematopoiesis was detected in 18.4% of people over 90 years of age. Ten-year myeloid malignancy probability ranged from 0.78% to 85% across risk strata. Of these cases, 88.4% were in the low-risk group [23][24]. Clones were common. Transformation was rare. This finding shows that mutation is a necessary condition. Mutation is not a sufficient condition. The peak of early-onset solid cancer moves with birth cohort. Each standard-deviation increase in biological age was associated with early-onset solid cancer risk, with an HR of 1.08 [29]. This is a readout of the adult branch. The adult-branch readout contrasts with the fixed peak of childhood cancer.

The disagreement between this paper and the two-step model is resolvable. The two-step model places the second hit in postnatal infection exposure. This paper places the second hit in correction failure and gate failure within the developmental window. Drivers of infection exposure should move the peak. The construction window should not move the peak. Both classes of variable can be measured in the same cohort.

4.3 Biological interpretation

Four termination mechanisms constitute the intrinsic gate. The four mechanisms are positional information, differentiation coupling, contact inhibition and programmed termination [18]–[22]. The four mechanisms are independent. Gate non-construction is a matter of degree and of different missing combinations. Gate non-construction does not equal the loss of one molecule. This explains why redundancy often absorbs a single-molecule knockout. This also explains why the molecular inventory of childhood cancer cannot directly give the entry state.

Membrane potential is the physical readout of the constructed state. Resting membrane potential of healthy human skeletal muscle changes by less than 1 mV between ages 1 and 99 [25]. Differentiation is a fall of membrane potential towards hyperpolarisation. Cancer cells are more depolarised

[26][27]. Human fetal neural stem cells are about − 30 mV [28]. The cell of origin in childhood cancer sits between the proliferative state and the differentiated state. This position can be read against the normal side of the same cell type.

Wilms tumour, neuroblastoma, retinoblastoma, hepatoblastoma and embryonal central nervous system tumour have different developmental windows. The five subtypes have different peak positions. The window is set by the maturation time of the cell of origin. The window is not set by the mutation rate.

4.4 Clinical implications

First, monitoring is timed to the developmental window. Children who carry a susceptible lesion or a known lineage risk are monitored according to their developmental window. Monitoring in these children is not distributed evenly by calendar age.

Second, the screening target is not a precancerous lesion. No multi-step precancerous sequence is available for screening in the general paediatric population. Screening for a precancerous lesion has no target on the paediatric branch. The alternative is to read whether construction is complete within the window.

Third, the intervention variables are clone-regression rate and gate-failure rate. The intervention variables are not mutation rate. The model simulation shows that a 50% rise in kR reduces cumulative clinical incidence by 24.4%–26.7%. The model simulation shows that a 50% fall in kG0 has an effect of 0%–21.1%. This difference indicates that raising clone regression within the window has a larger effect than lowering gate failure alone.

4.5 Identifiability and negative results

The parameters are not unique. Five parameters were estimated from five age-group rates. Different parameter sets can produce similar curves. aG and σG were set from developmental literature. aG and σG did not enter the fit. Peak groups were sensitive to a 0.5-year upward shift in aG. Neuroblastoma and retinoblastoma moved to 1–4 years under this shift. Peak order was insensitive to σG.

The single-branch model could not reproduce the early peak. The single-branch AIC was higher than the latent-variable AIC in all five subtypes. The ΔAIC for hepatoblastoma was only 3.1. Hepatoblastoma was the weakest of the five subtypes.

The peak group for embryonal central nervous system tumour was predicted one age group late. The observed rates at < 1 year and 1–4 years differed by 1.7 per million. The continuous peak of the model lay at 1–2 years. This difference did not change peak order. This difference also did not change the conclusion that childhood cancer follows developmental windows.

The clone-to-clinical ratio came from two data types: model output and autopsy series. Selection bias, age composition and pathology criteria in autopsy series can affect the ratio. The model gives an expected value under the same kinetics. The model is not a fit to one autopsy series.

4.6 Conclusion

The origin of childhood cancer is persistent fetal developmental-sequence deviation, growth-phase correction failure and gate failure. The rate of mutation accumulation cannot produce the early peak of childhood cancer. The single-branch mutation model cannot reproduce the peak groups of the five subtypes. The five-state latent-variable model reproduces age-specific incidence and peak order. The model gives a clone-to-clinical ratio of 37.6–72. The model turns clone-regression rate and gate-failure rate within the window into simulable intervention variables.

5 Conclusion

The entry point of childhood cancer is not the rate of mutation accumulation. The entry point of childhood cancer is persistent fetal developmental-sequence deviation and growth-phase correction failure. The clinical tumour is the terminal readout of this path. The clone-regression rate within the window is an earlier intervention variable than the mutation rate.

Declarations

Ethics approval and consent to participate

This study used published, de-identified registry data and literature. This study involved no new human participants, human tissue or animals.

Data and material availability

SEER data are publicly available. Model code, parameter tables and result files are provided with the paper.

Conflicts of interest

The authors declare no conflicts of interest.

Funding

This study received no funding.

Author contributions

Qiao-Xuan Wei: conceptual construction, model design, data analysis, first draft and revision. Yu-Ting Ma: clinical interpretation, evidence verification and revision. Gui-Juan Zhu: data checking, table and figure checking and revision.

References

1. Schwartz LF et al (2026) Advances and disparities in survival of childhood and adolescent cancers. J Natl Cancer Inst 118(10):1919–1932 PMID: 42149832
2. Youlden DR et al (2023) Childhood cancer survival and avoided deaths in Australia, 1983–2016. Paediatr Perinat Epidemiol 37(1):81–91 PMID: 35672573
3. Chang TC et al (2024) Genomic determinants of outcome in acute lymphoblastic leukemia. J Clin Oncol 42(29):3491–3503 PMID: 39121442
4. GBD 2023 Childhood Cancer Collaborators (2026) Global burden of cancer in children and adolescents aged 0–19 years, 1990–2023. Lancet 407(10536):1360–1373 PMID: 41936367
5. Greaves M (2018) A causal mechanism for childhood acute lymphoblastic leukaemia. Nat Rev Cancer 18(8):471–484 PMID: 29784935
6. Ford AM et al (1993) In utero rearrangements in the trithorax-related oncogene in infant leukaemias. Nature 363(6427):358–360 PMID: 8497319
7. Brown EG et al (2023) Unwrapping nephrogenic rests and nephroblastomatosis for pediatric surgeons. J Pediatr Surg 58(11):2128–2134 PMID: 37625940
8. Labuhn M et al (2019) Mechanisms of progression of myeloid preleukemia to transformed myeloid leukemia in children with Down syndrome. Cancer Cell 36(2):123–138 PMID: 31303423
9. Singh HP et al (2018) Developmental stage-specific proliferation and retinoblastoma genesis in RB-deficient human but not mouse cone precursors. Proc Natl Acad Sci USA. ;115(40):E9391-E9400. PMID: 30213853
10. Ye K et al (2026) Longitudinal analysis of retinal cell state transitions in RB1-deficient retinal organoids. Cell Death Dis. ;17(1):34. PMID: 41535675
11. Sweet-Cordero EA et al (2019) The genomic landscape of pediatric cancers. Science 363(6432):1170–1175 PMID: 30872516
12. Jessa S et al (2019) Stalled developmental programs at the root of pediatric brain tumors. Nat Genet 51(12):1702–1713 PMID: 31768071
13. Steliarova-Foucher E et al (2017) International incidence of childhood cancer, 2001-10. Lancet Oncol 18(6):719–731 PMID: 28410997
14. National Cancer Institute. SEER, Cancer Statistics, Review (2015) 1975–2012, Table 29.1: Age-adjusted and age-specific SEER cancer incidence rates, 2008–2012, by ICCC group and age. Bethesda, MD: National Cancer Institute; https://seer.cancer.gov/csr/1975_2012/
15. Conces MR et al (2026) Pathology of peripheral neuroblastic tumors. Diagn Pathol. ;21(1):22. PMID: 41612478
16. Vuononvirta R et al (2008) Perilobar nephrogenic rests are nonobligate molecular genetic precursor lesions of insulin-like growth factor-II-associated Wilms tumors. Clin Cancer Res 14(23):7635–7644 PMID: 19047088
17. Boucher AC et al (2021) Clinical and biological aspects of myeloid leukemia in Down syndrome. Leukemia 35(12):3352–3360 PMID: 34518645
18. Takahashi A (2024) The discovery of oncogene-induced senescence. Nat Rev Mol Cell Biol 25(12):951 PMID: 39384956
19. Edwards-Hicks J et al (2022) MYC sensitises cells to apoptosis by driving energetic demand. Nat Commun. ;13:4674. PMID: 35945217
20. Vafa O et al (2002) c-Myc can induce DNA damage, increase reactive oxygen species, and mitigate p53 function. Mol Cell 9(5):1031–1044 PMID: 12049739
21. Gaillard H, Garcia-Muse T, Aguilera A (2015) Replication stress and cancer. Nat Rev Cancer 15(5):276–289 PMID: 25907220
22. Wu K et al (2026) Targeting FSP1 triggers ferroptosis in lung cancer. Nature 649(8096):487–495 PMID: 41193800
23. Jaiswal S et al (2014) Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med 371(26):2488–2498 PMID: 25426837
24. Weeks LD et al (2023) Prediction of risk for myeloid malignancy in clonal hematopoiesis. NEJM Evid PMID:37483562
25. Forsberg AM et al (1997) Resting membrane potential of skeletal muscle calculated from plasma and muscle electrolyte and water contents. Clin Sci (Lond) 92(4):391–396 PMID: 9176039
26. Bautista W et al (2017) Cancer stem cells are depolarized relative to normal stem cells derived from human livers. Ann Hepatol 16(2):297–303 PMID: 28233753
27. Bautista W et al (2014) Membrane potential differences and GABAA receptor expression in hepatic tumor and non-tumor stem cells. Can J Physiol Pharmacol 92(1):85–91 PMID: 24383877
28. De Zio R et al (2026) A conserved bioelectrical signature defines subventricular zone-derived human fetal neural stem cells and tracks their differentiation state. Front Cell Dev Biol. ;14:1774119. PMID: 41890905

29. Tian R et al (2026) Biological aging and generational shifts in early-onset cancer risk. Nat Med 32(8):2983–2990 PMID: 42332142
30. Zhao H et al (2008) Post-transcriptional down-regulation of Atoh1/Math1 by bone morphogenic proteins suppresses medulloblastoma development. Genes Dev 22(6):722–727 PMID: 18347090
31. Filipponi D et al (2022) Inhibiting ALK2/ALK3 signaling to differentiate and chemo-sensitize medulloblastoma. Cancers 14(9):2095 PMID: 35565225
32. Kool M et al (2012) Molecular subgroups of medulloblastoma: an international meta-analysis. Acta Neuropathol 123(4):473–484 PMID: 22358457
33. Northcott PA et al (2017) The whole-genome landscape of medulloblastoma subtypes. Nature 547(7663):311–317 PMID: 28726821
34. Papaemmanuil E et al (2014) RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. Nat Genet 46(2):116–125 PMID: 24413735
35. Rivera MN, Haber DA (2005) Wilms' tumour: connecting tumorigenesis and organ development in the kidney. Nat Rev Cancer 5(9):699–712 PMID: 16110318
36. Garaventa A et al (2021) Randomized trial of two induction therapy regimens for high-risk neuroblastoma: HR-NBL1.5. J Clin Oncol 39(23):2552–2563 PMID: 34152804
37. Vora A et al (2006) Toxicity and efficacy of 6-thioguanine versus 6-mercaptopurine in childhood lymphoblastic leukaemia. Lancet 368(9544):1339–1348 PMID: 17046466
38. Schrappe M et al (2018) Reduced-intensity delayed intensification in standard-risk pediatric acute lymphoblastic leukemia. J Clin Oncol 36(3):244–253 PMID: 29148893
39. Oeffinger KC et al (2006) Chronic health conditions in adult survivors of childhood cancer. N Engl J Med 355(15):1572–1582 PMID: 17035650
40. Bhakta N et al (2017) The cumulative burden of surviving childhood cancer: St Jude Lifetime Cohort. Lancet 390(10112):2569–2582 PMID: 28890157
41. Dixon SB et al (2023) Specific causes of excess late mortality among survivors of childhood cancer. Lancet 401(10386):1447–1457 PMID: 37030315

Figures

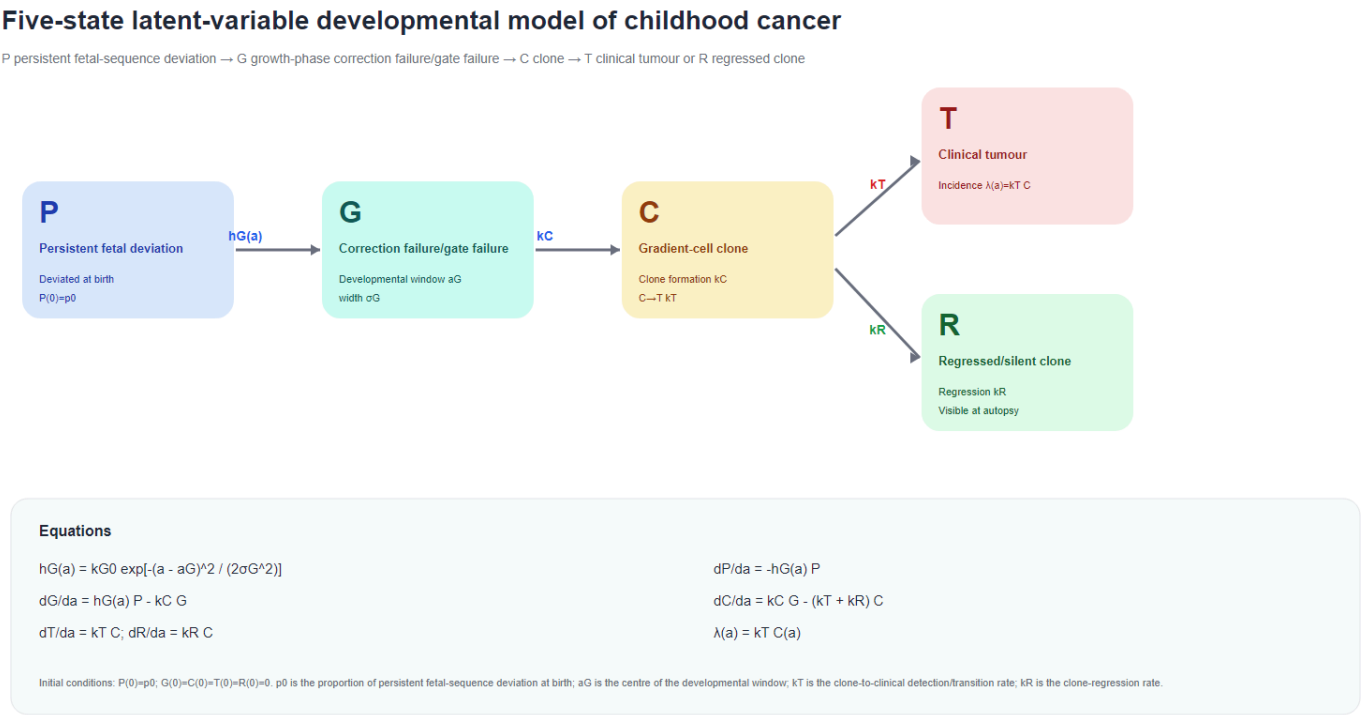


Figure 1

The five-state latent-variable developmental model of childhood cancer. P is persistent fetal developmental-sequence deviation. G is failure of growth-phase epigenetic correction and gate failure. C is a gradient-cell clone. T is a clinical tumour. R is a regressed or silent clone. hG(a) is the P→G transition intensity within the developmental window. kC is the G→C intensity. kT is the C→T intensity. kR is the C→R intensity. Incidence λ(a) equals kT multiplied by C(a).

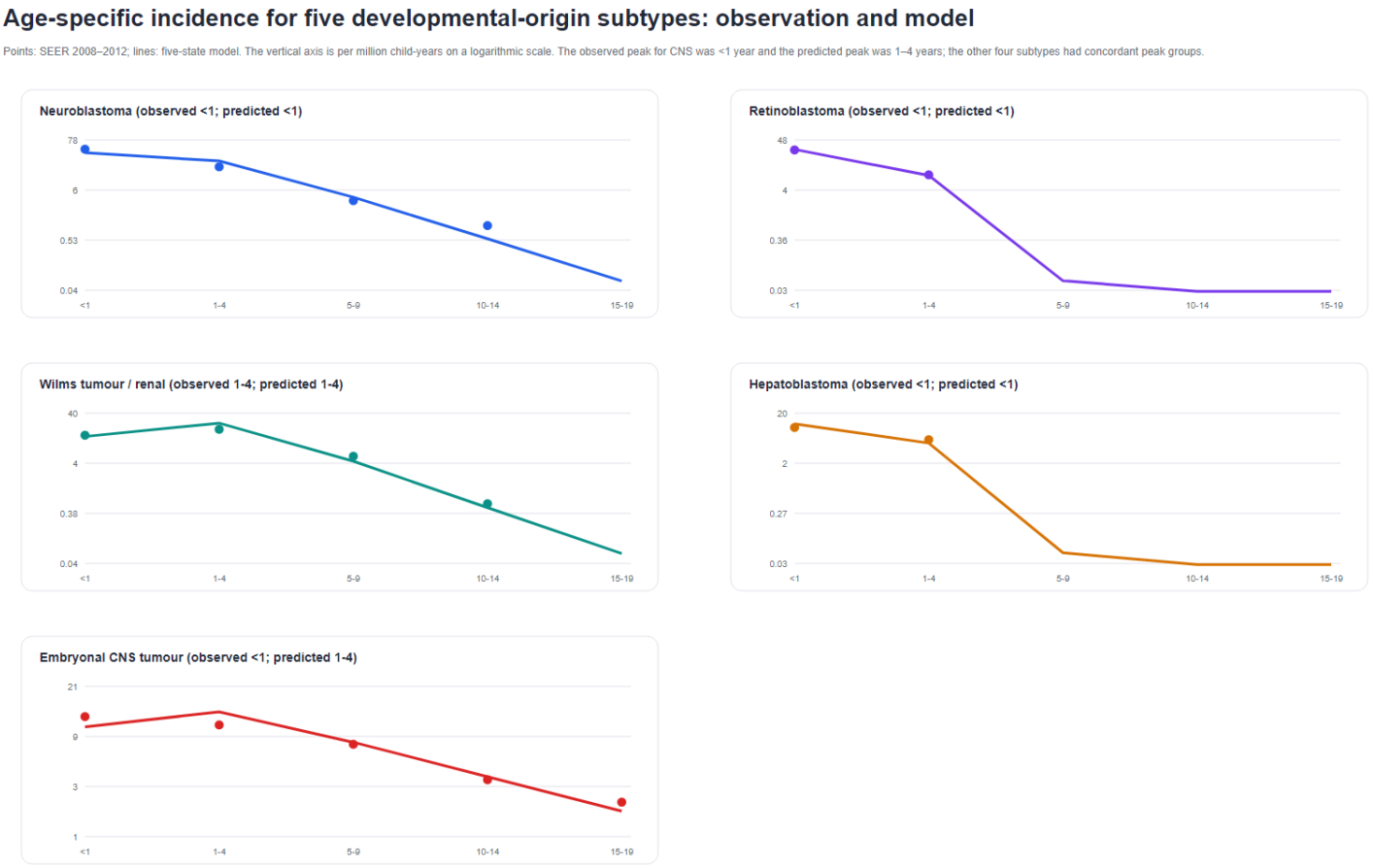


Figure 2

Age-specific incidence for five developmental-origin subtypes: observation and model. Points are SEER 2008–2012. Lines are the five-state model. The vertical axis is per million child-years on a logarithmic scale. The observed peak for embryonal central nervous system tumour was <1 year. The predicted peak for this subtype was 1–4 years. The other four subtypes had concordant peak groups.

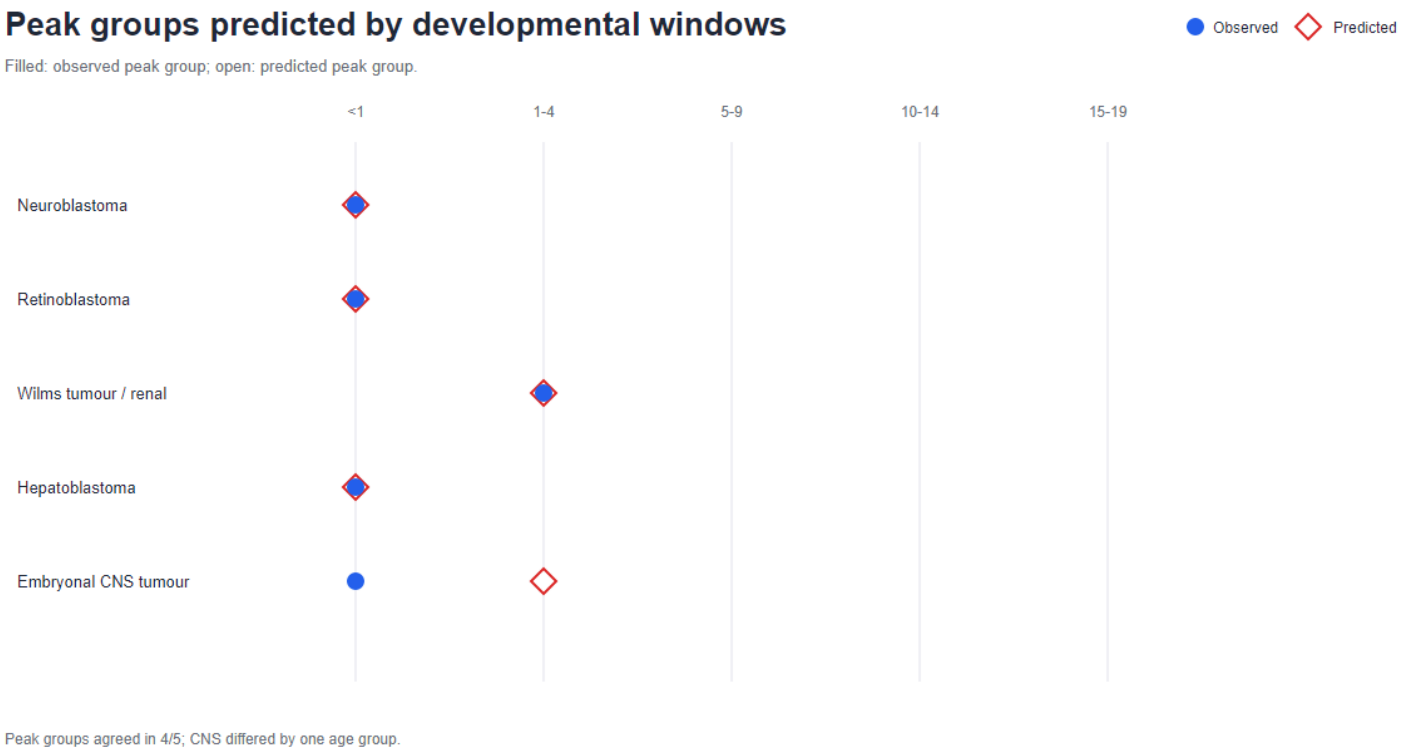


Figure 3

Peak groups predicted by developmental windows. Filled circles are observed peak groups. Open diamonds are predicted peak groups. Peak groups agreed in four of five subtypes. Embryonal central nervous system tumour differed by one age group.

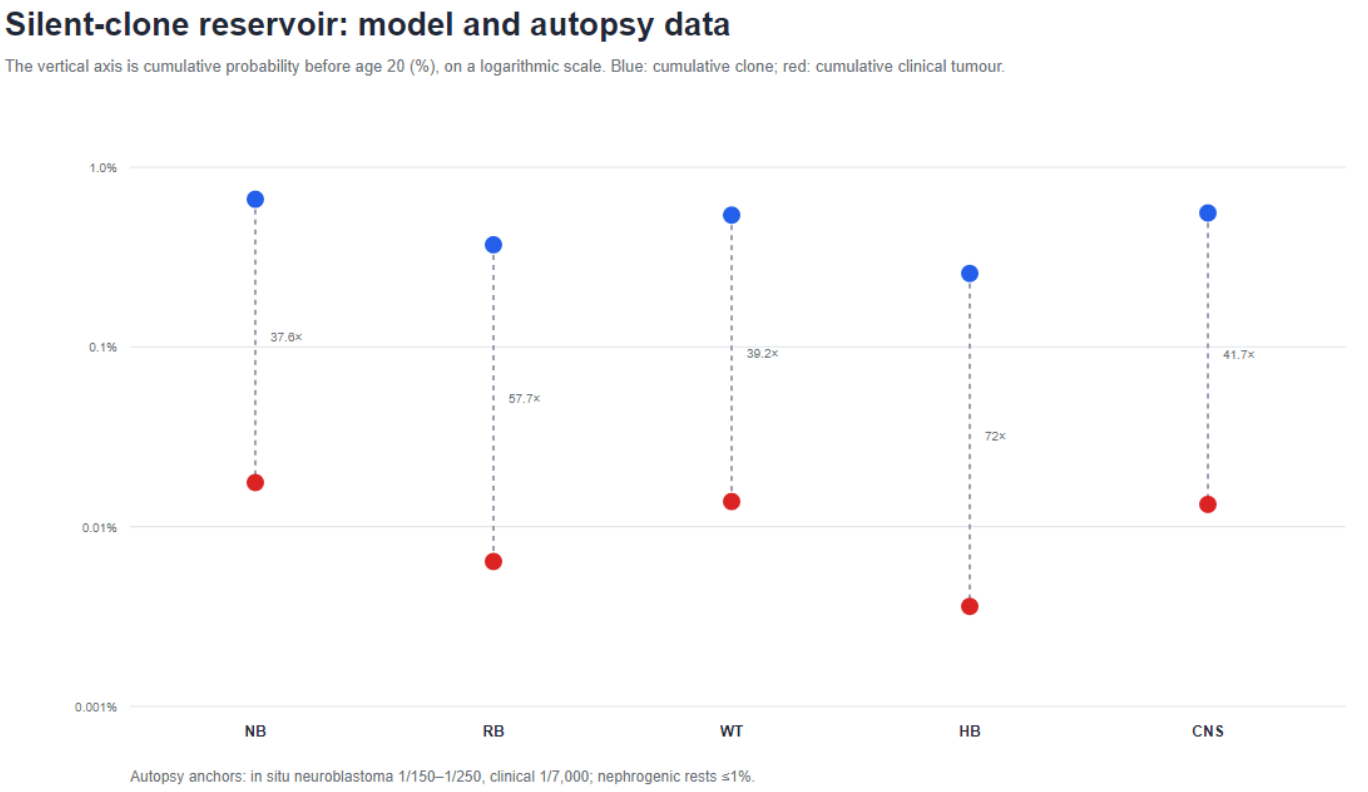


Figure 4

The silent-clone reservoir: model and autopsy data. The vertical axis is cumulative probability before age 20 (%), on a logarithmic scale. Blue points are cumulative clone. Red points are cumulative clinical tumour. Dashed lines connect the two readouts for the same subtype. Numbers are clone-to-clinical ratios. Autopsy anchors are in situ neuroblastoma 1/150–1/250, clinical neuroblastoma 1/7,000 and nephrogenic rests ≤1%.

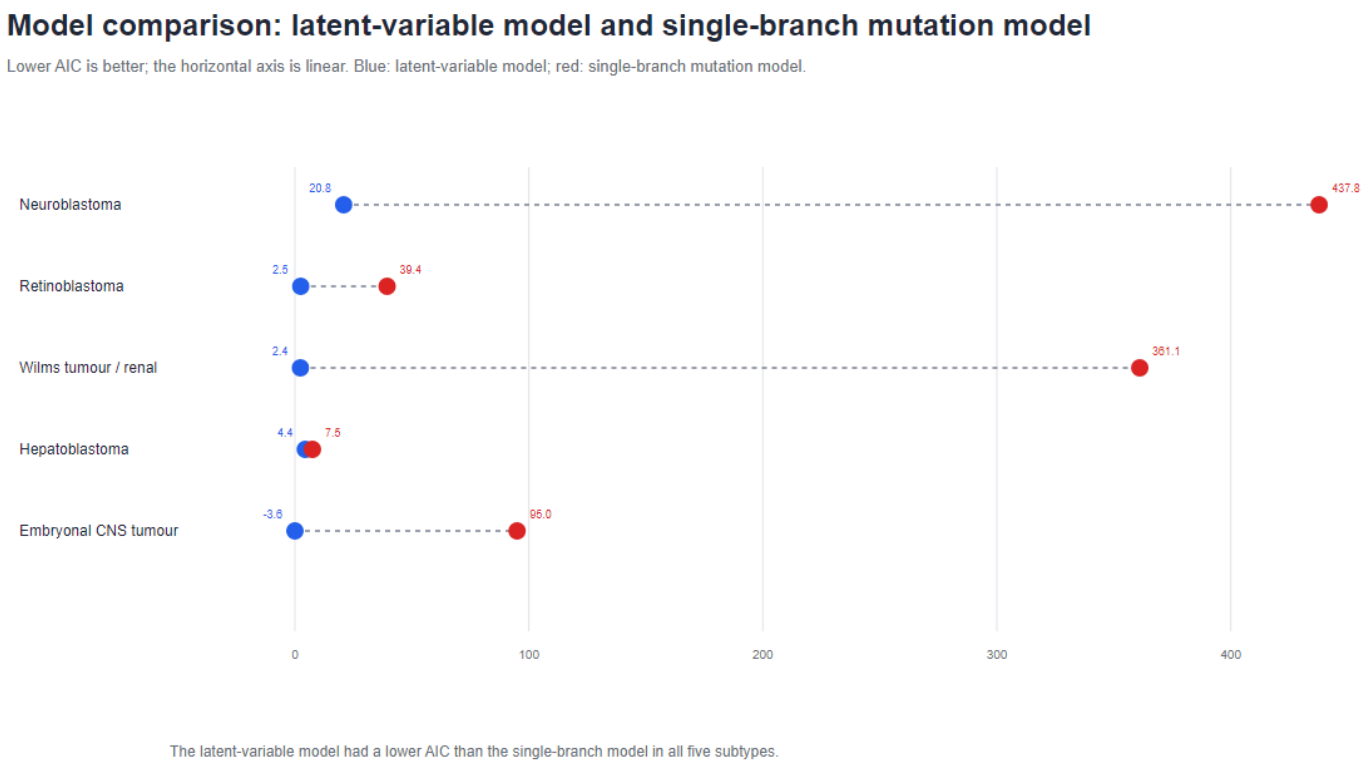


Figure 5

Model comparison: latent-variable model and single-branch mutation model. The horizontal axis is AIC. Blue points are the latent-variable model. Red points are the single-branch model. The latent-variable model had a lower AIC in all five subtypes.

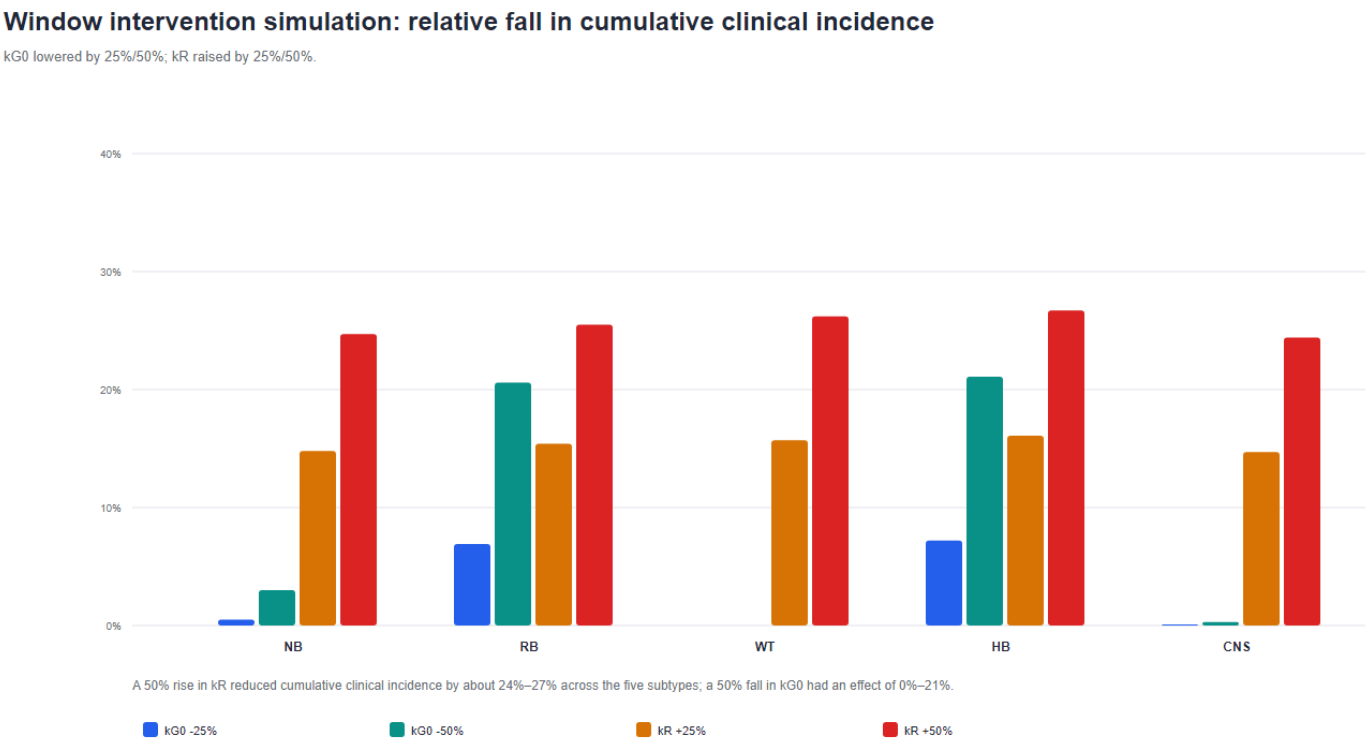


Figure 6

Window intervention simulation: relative fall in cumulative clinical incidence. kG0 was lowered by 25% and 50%. kR was raised by 25% and 50%. A 50% rise in kR reduced cumulative clinical incidence by about 24%–27% across the five subtypes. A 50% fall in kG0 had an effect of 0%–21%.