
SCOR
The Art & Science of Risk

Accelerated Underwriting and, Underwriting with Partial Information
IDSC 2024

Jayant Apte, Antoine Ly

Data Analytics Solutions, Scor

SCOR
The Art & Science of Risk

Accelerated Underwriting and, Underwriting with Partial Information
High Level Positioning

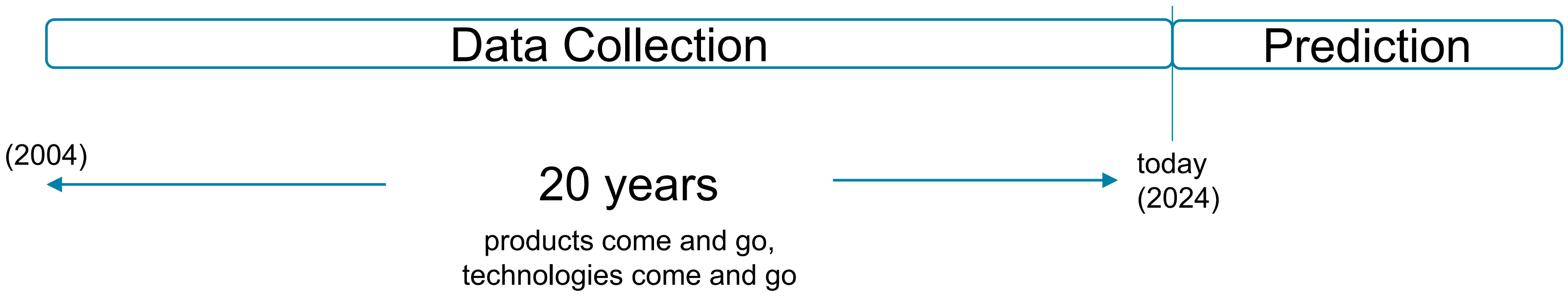
- Algorithmic Underwriting refers to the use of computational algorithms, external data sources, and Big Data solutions to inform an underwriting decision
    - We are interested in the use of machine learning techniques for predictive modeling to improve some KPI associated with life insurance underwriting
        - We are interested in using historical, policy level, life insurance claims (events) data
            - Specifically, we are talking about the kind of data that is generated by life insurance underwriting systems in United States over the span of last 20 years

SCOR
The Art & Science of Risk

Accelerated Underwriting and, Underwriting with Partial Information

High Level Positioning

- Algorithmic Underwriting refers to the use of computational algorithms, external data sources, and Big Data solutions to inform an underwriting decision
  - We are interested in the use of machine learning techniques for predictive modeling to improve some KPI associated with life insurance underwriting
    - We are interested in using historical, policy level, life insurance claims (events) data
      - Specifically, we are talking about the kind of data that is generated by life insurance underwriting systems in United States over the span of last 20 years



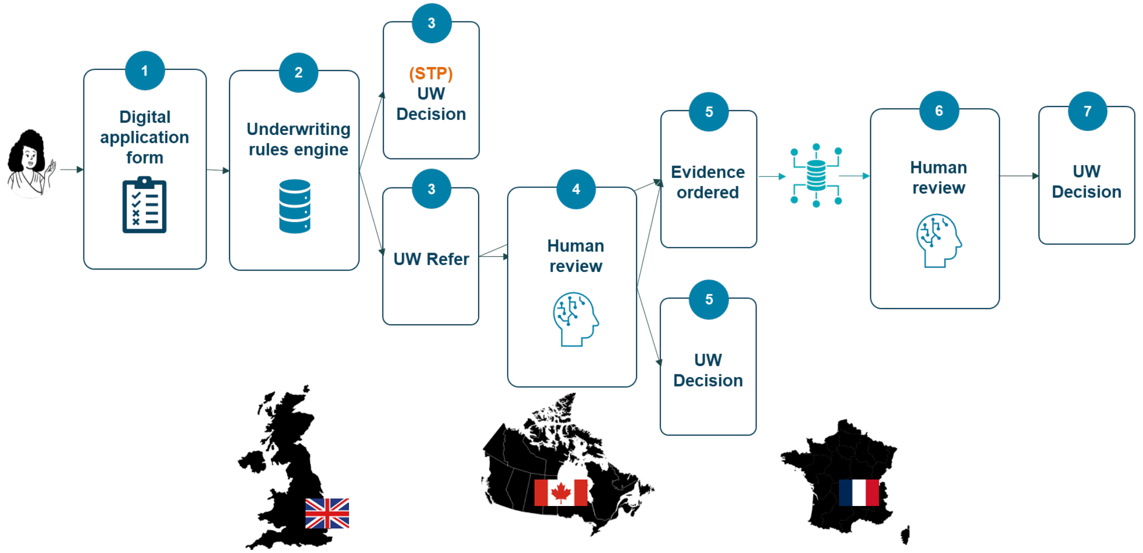
Accelerated Underwriting and, Underwriting with Partial Information
High Level Positioning

- Algorithmic Underwriting refers to the use of computational algorithms, external data sources, and Big Data solutions to inform an underwriting decision
  - We are interested in the use of machine learning techniques for predictive modeling to improve some KPI associated with life insurance underwriting
    - We are interested in using historical, policy level, life insurance claims (events) data
      - Specifically, we are talking about the kind of data that is generated by life insurance underwriting systems in United States over the span of last 20 years
        - We are interested in creating a high-level conceptual framework for insurers to use when thinking about historical claims data and its applications
          - We want to understand and solve various challenges one may encounter when using such data to learn machine learning models

Global Context

Life underwriting journey in EU and several other parts of the world

Majority of applicants are straight-through processed but ~30% need human review

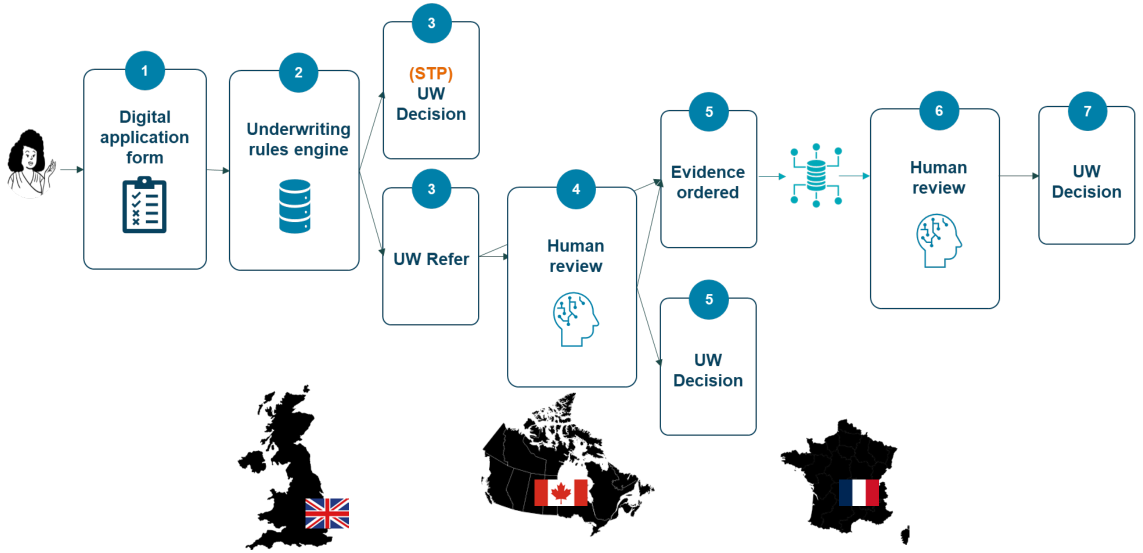


- Several countries in the EU and elsewhere in the world have a straightforward underwriting process
- A large fraction of applicants (60-90%) go through "straight through processing" (STP) which means they get policies in real time
- In the UK, the STP constitutes 70-80% of cases while for Ireland it is 60%
- In the UK, of the cases that don't go through automated process, only about 10% get extra data sources such as EHR while rest are assessed by an underwriter manually
- In France, no medical questions are allowed under 200k (Loi Lemaine)

Global Context

Life underwriting journey in EU and several other parts of the world

Majority of applicants are straight-through processed but ~30% need human review



- Several countries in the EU and elsewhere in the world have a straightforward underwriting process
- A large fraction of applicants (60-90%) go through “straight through processing” (STP) which means they get policies in real time
- In the UK, the STP constitutes 70-80% of cases while for Ireland it is 60%
- In the UK, of the cases that don’t go through automated process, only about 10% get extra data sources such as EHR while rest are assessed by an underwriter manually
- In France, no medical questions are allowed under 200k (Loi Lemaine)

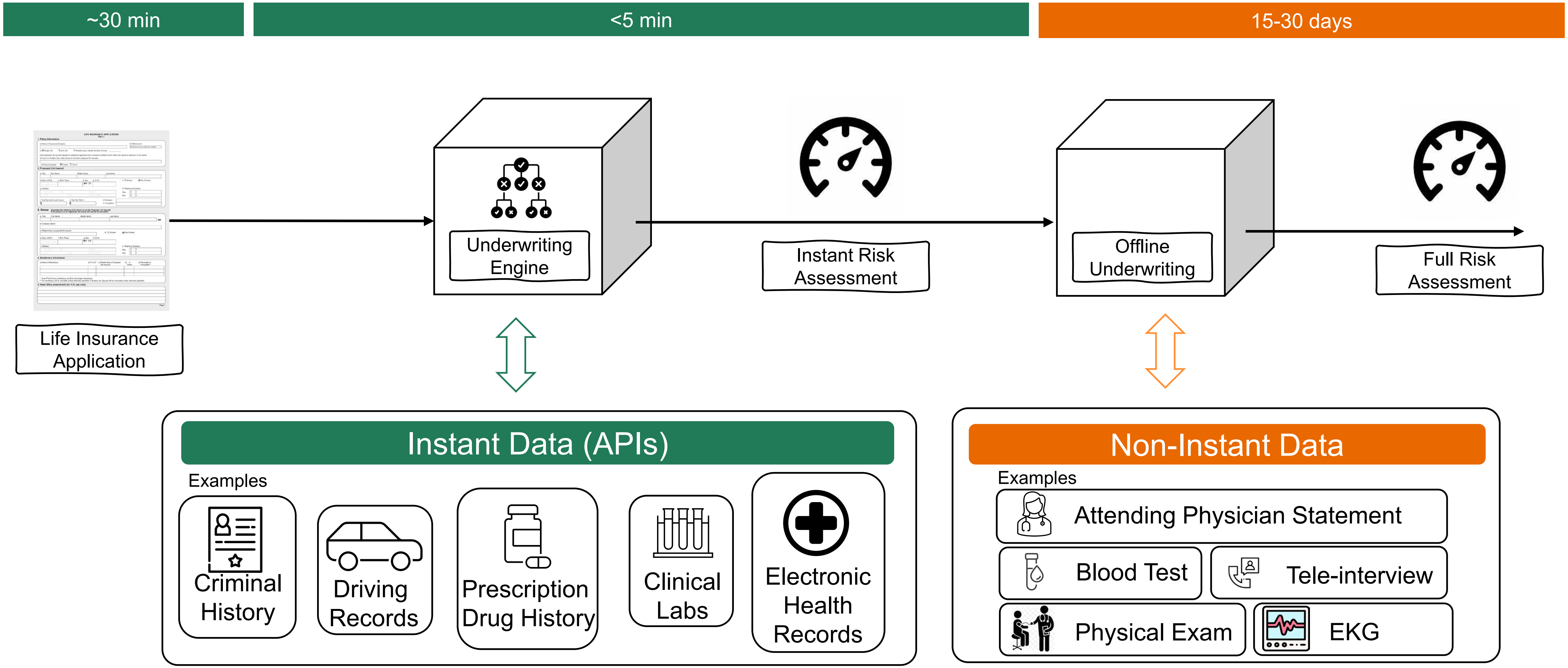
- Insurers can order third party vendor data via web APIs
- >10 types of data can be ordered
- Tens of data vendors providing on-demand data

Accelerated Underwriting and, Underwriting with Partial Information
Agenda

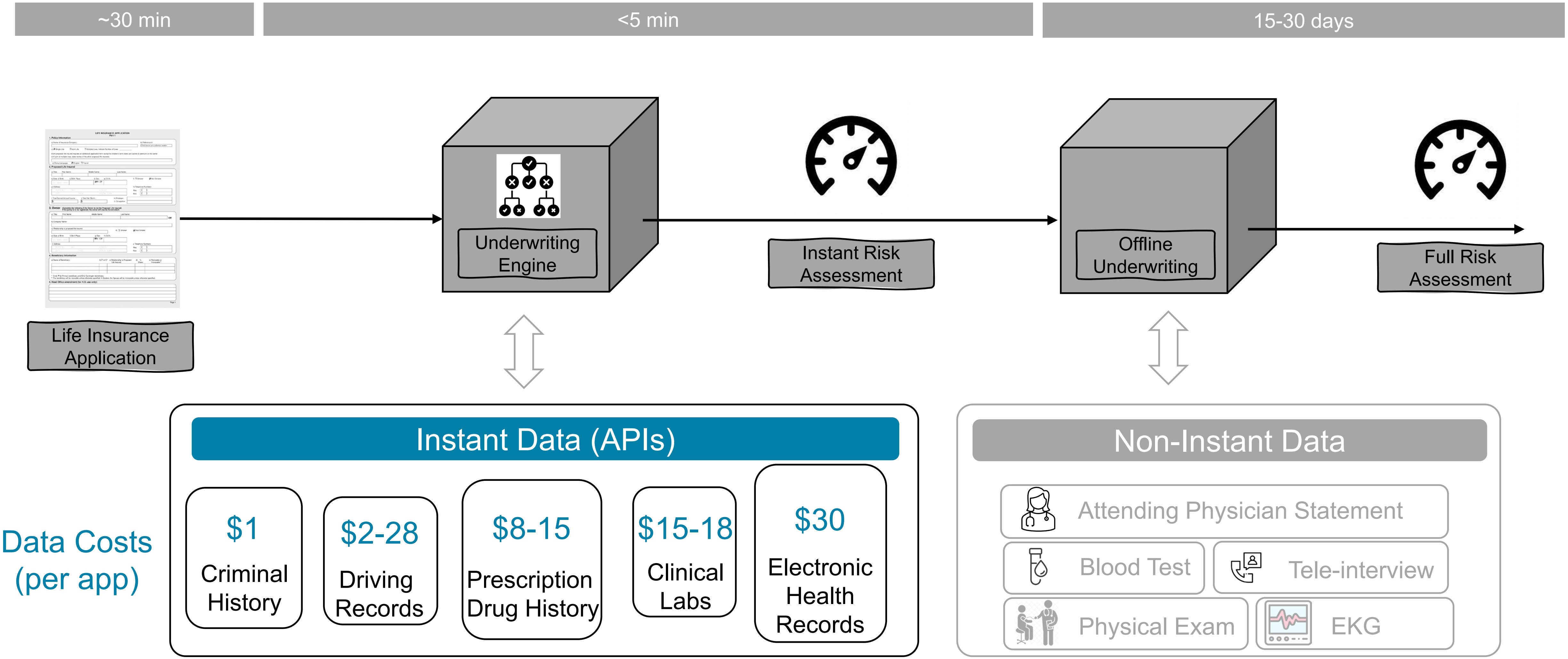
- Introduction to Accelerated Underwriting (AU) works today
- Introduction to evidence waiver models
- How to use life insurance claims data to learn waiver models
- Challenges and methods

Algorithmic Underwriting & Accelerated Underwriting (AU) Programs in US Life Market

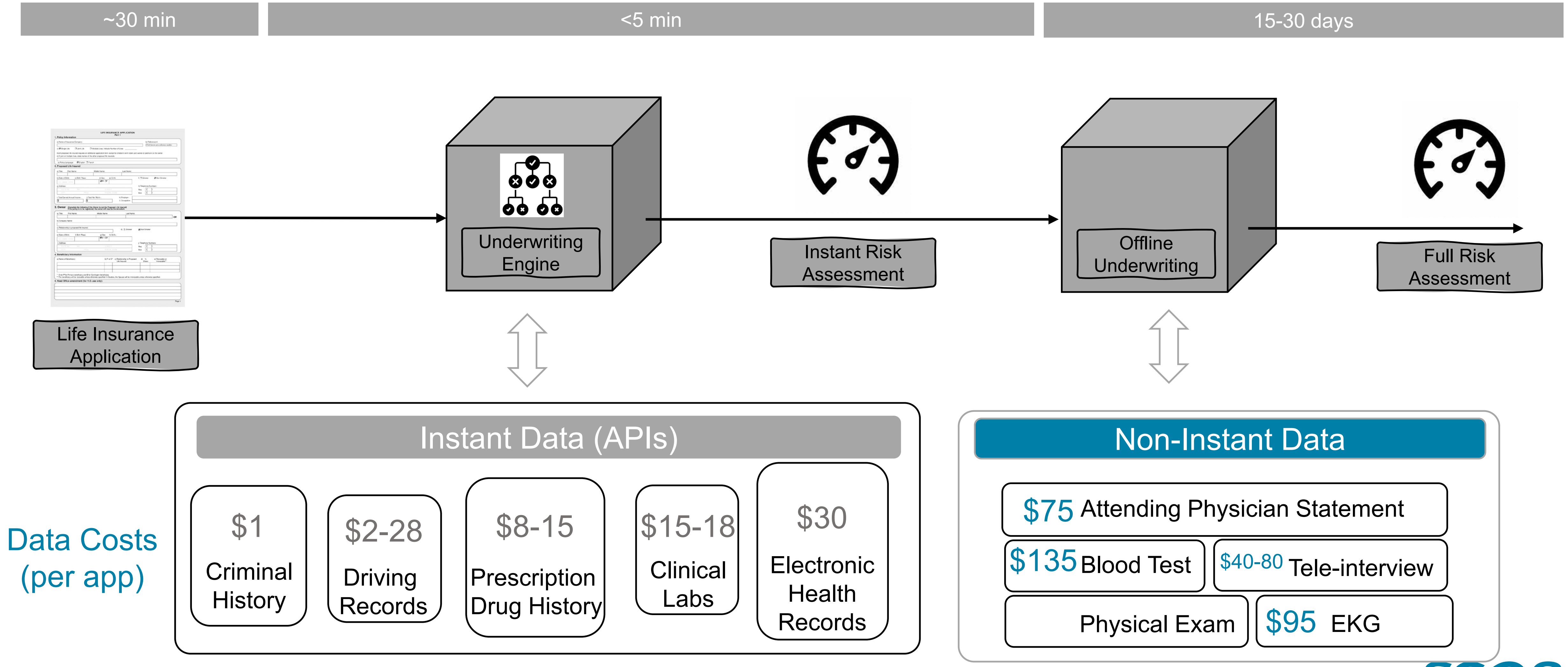
Algorithmic Underwriting (AUW) in US Life Insurance



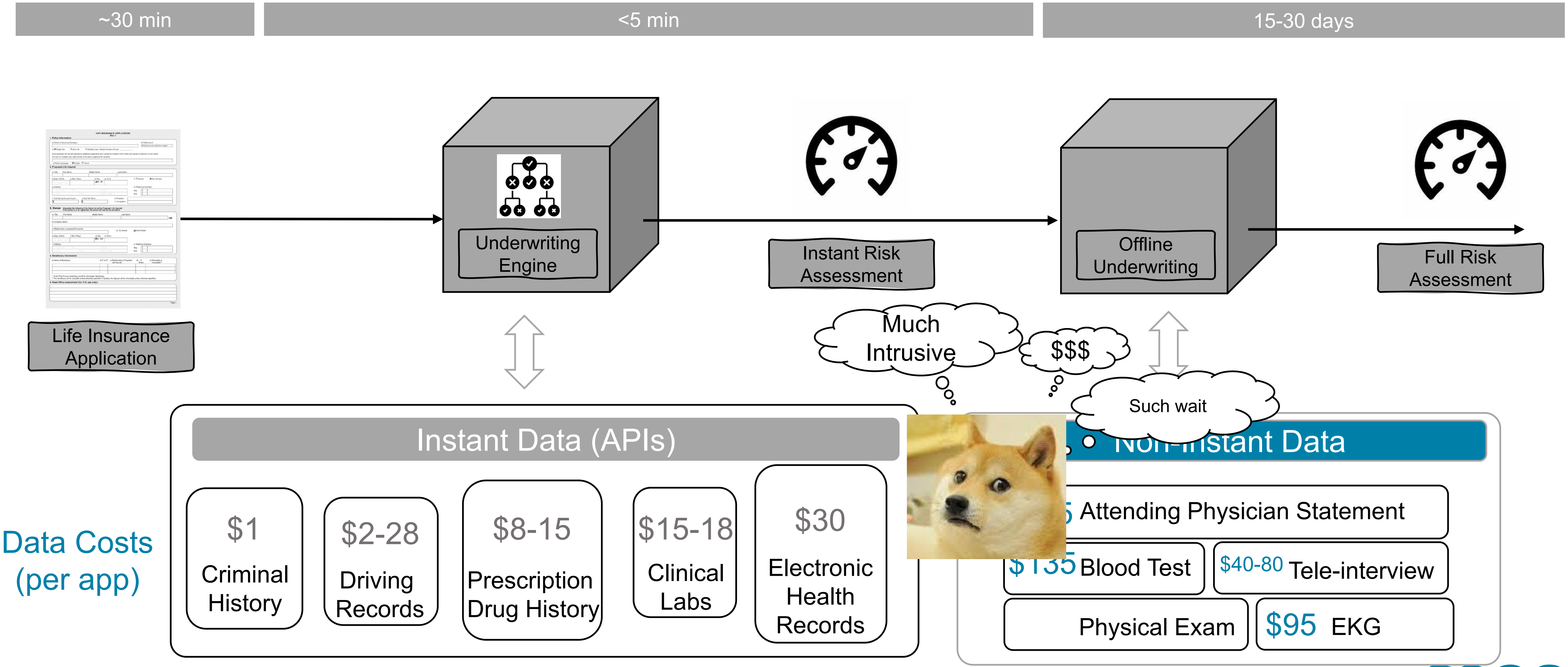
Algorithmic Underwriting (AUW) in US Life Insurance



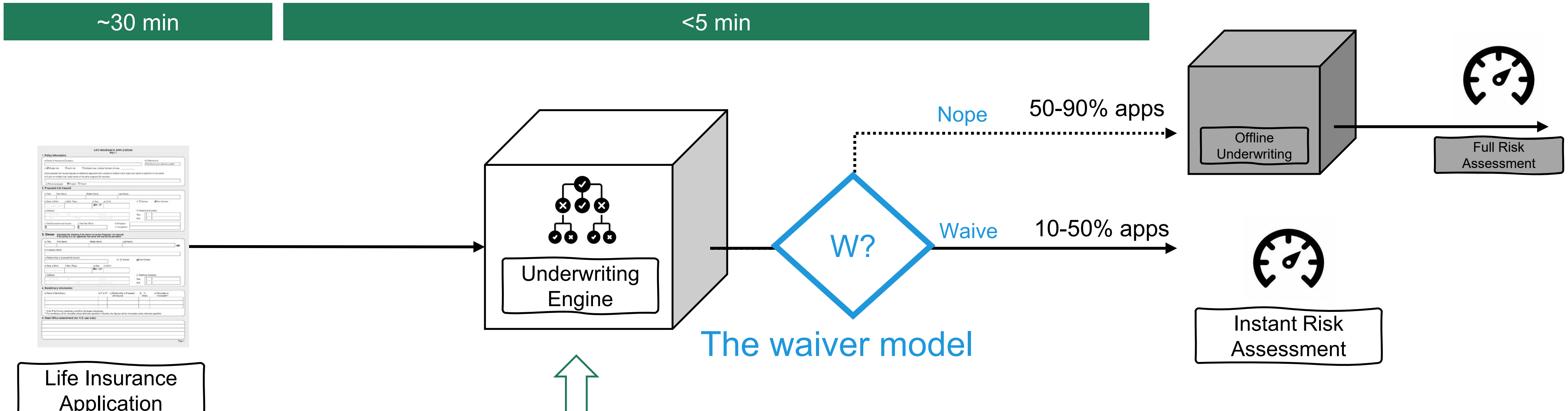
Algorithmic Underwriting (AUW) in US Life Insurance

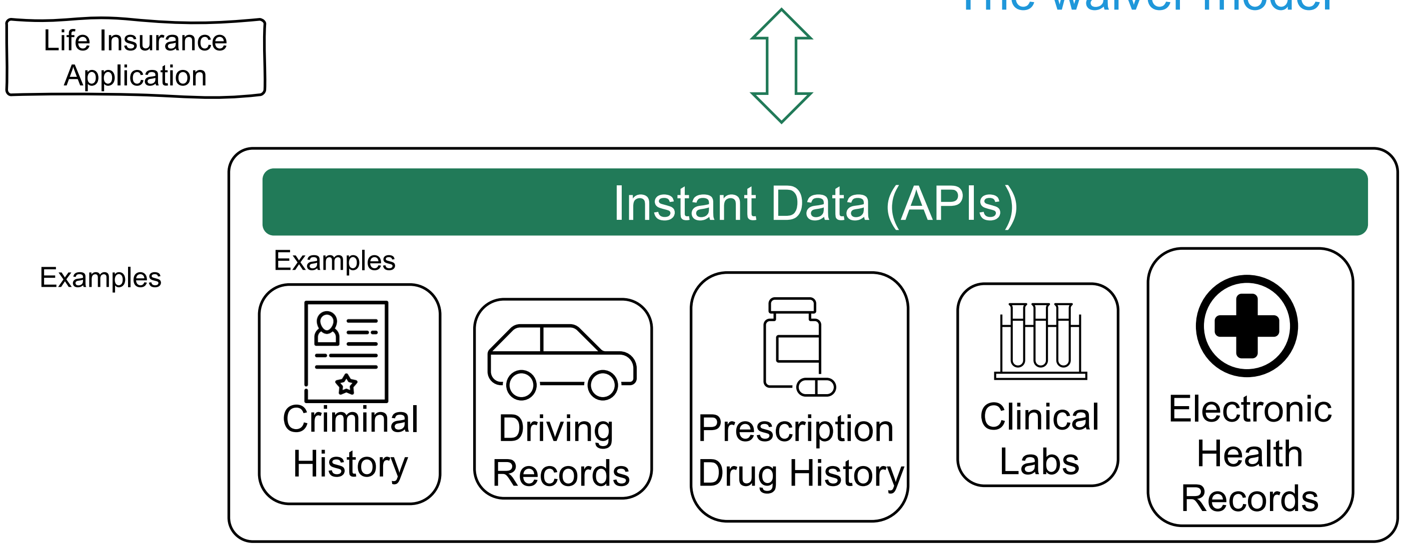


Algorithmic Underwriting (AUW) in US Life Insurance



Accelerated Underwriting (AU)

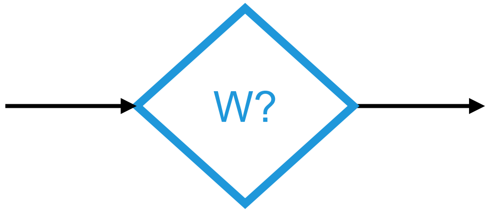




An “AU Program” is a program used for life insurance products where an applicant can have certain underwriting requirements waived, such as forgoing insurance fluid requirements and a paramedical exam, if they meet certain qualifications, typically determined by an algorithm used for this purpose. This algorithm will also typically determine the risk class the applicant will be offered.

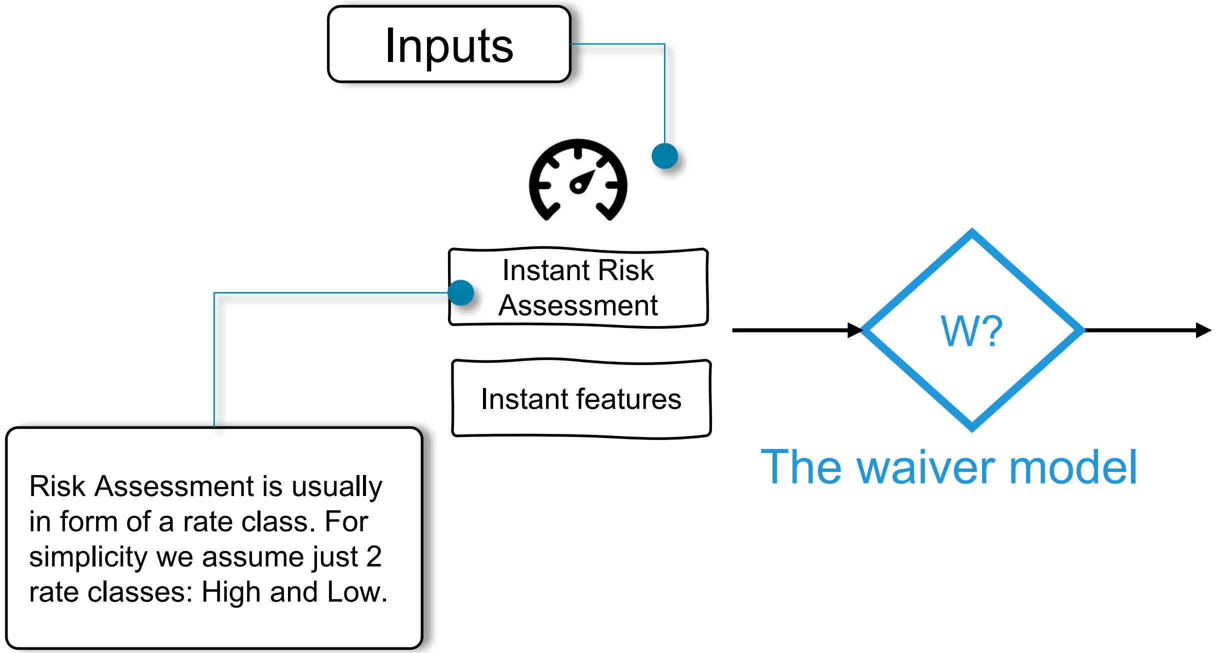
Klein A., Li J. (2023). 2022 Accelerated Underwriting Practices Survey Report. The Society of Actuaries Research Institute

Zoom on the waiver model

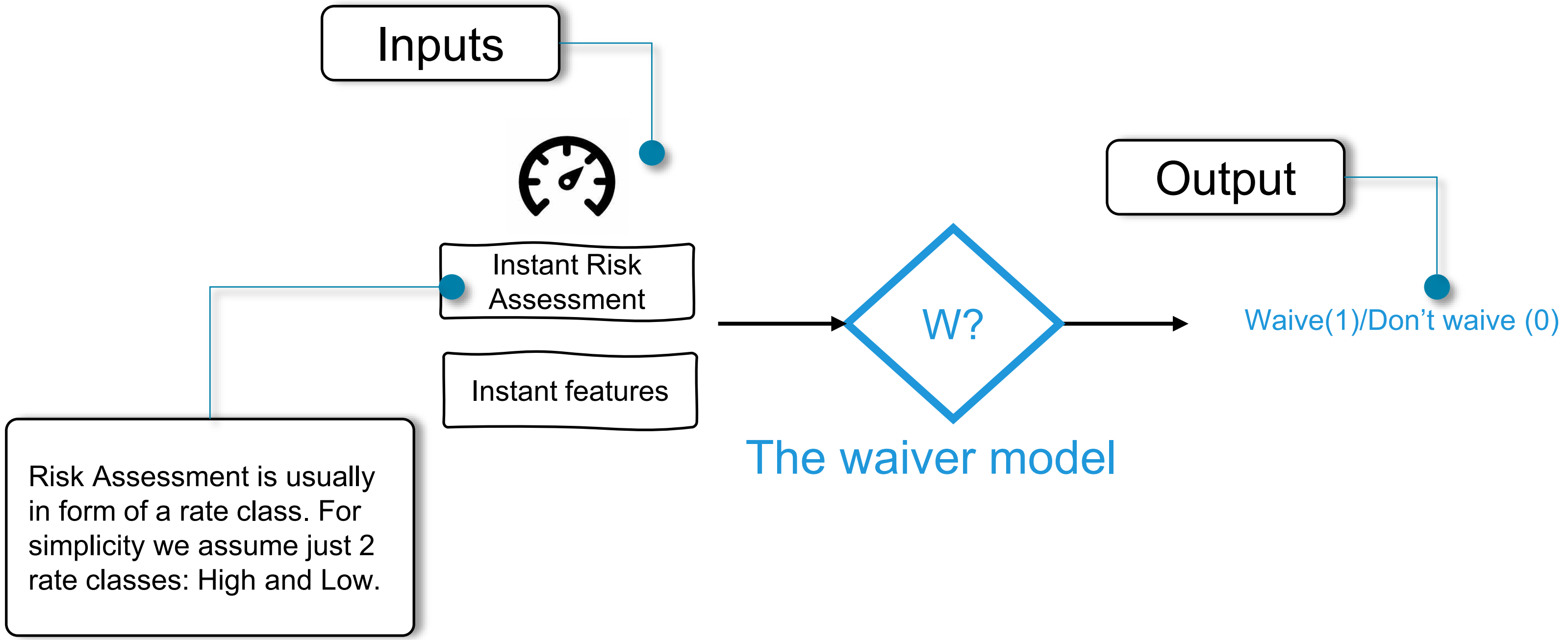


The waiver model

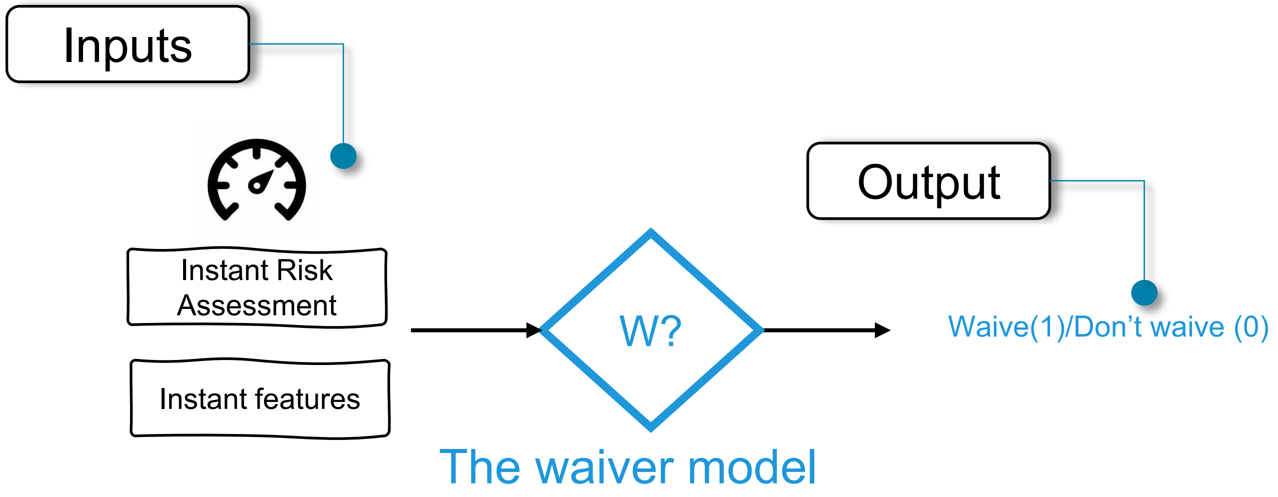
Zoom on the waiver model



Zoom on the waiver model



Zoom on the waiver model: performance metrics



These applicants will pay less premium due to faulty risk assessment. This makes the product vulnerable to anti-selection.

W? Risk Assessment

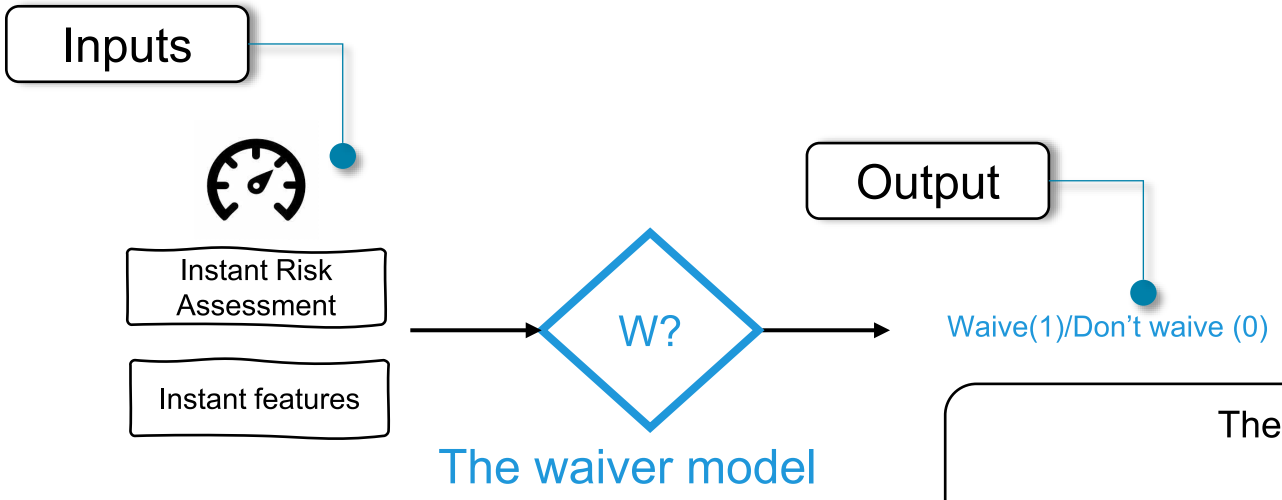
These applicants may feel hard done by the risk assessment they receive and will consider shopping elsewhere

Full Risk Assessment

	Low	High
Low	x	
High		x

SCOR The Art & Science of Risk

Zoom on the waiver model: performance metrics



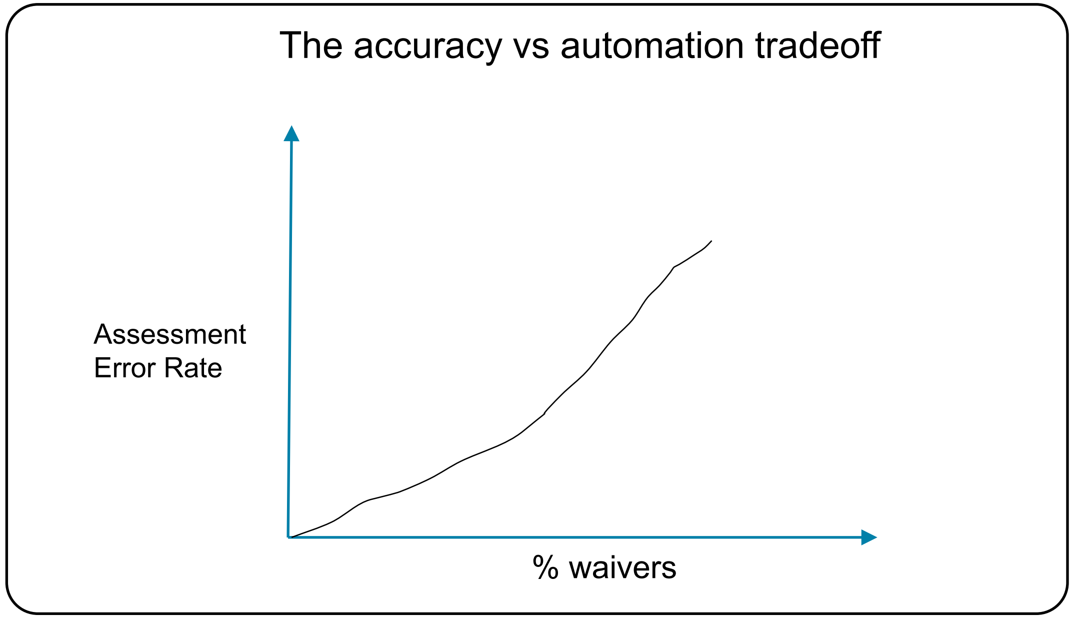
These applicants will pay less premium due to faulty risk assessment. This makes the product vulnerable to anti-selection.

W? Risk Assessment

These applicants may feel hard done by the risk assessment they receive and will consider shopping elsewhere

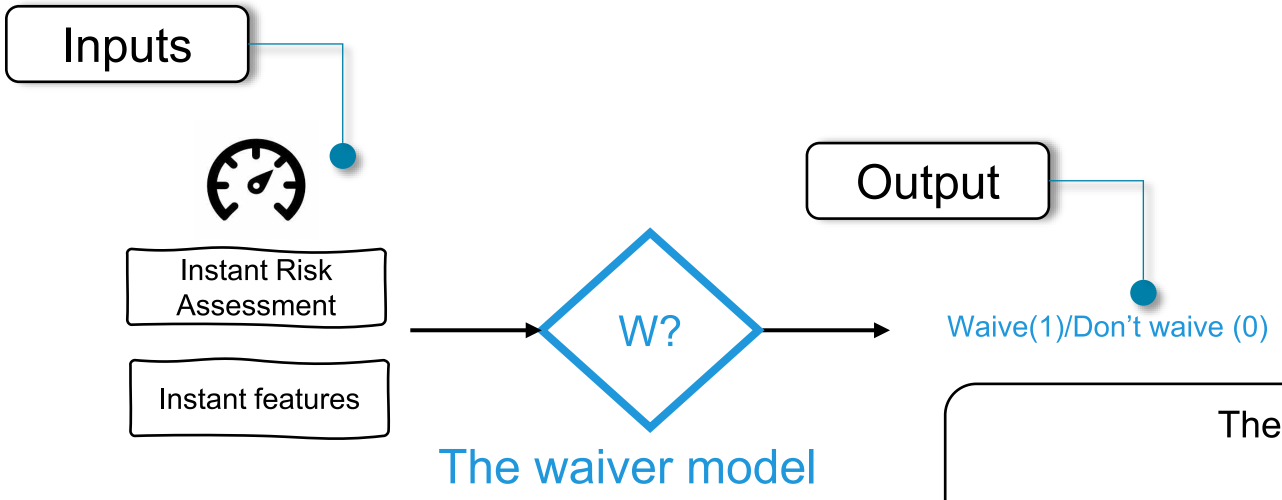
Full Risk Assessment

	Low	High
Low	x	
High		x



SCOR
The Art & Science of Risk

Zoom on the waiver model: performance metrics

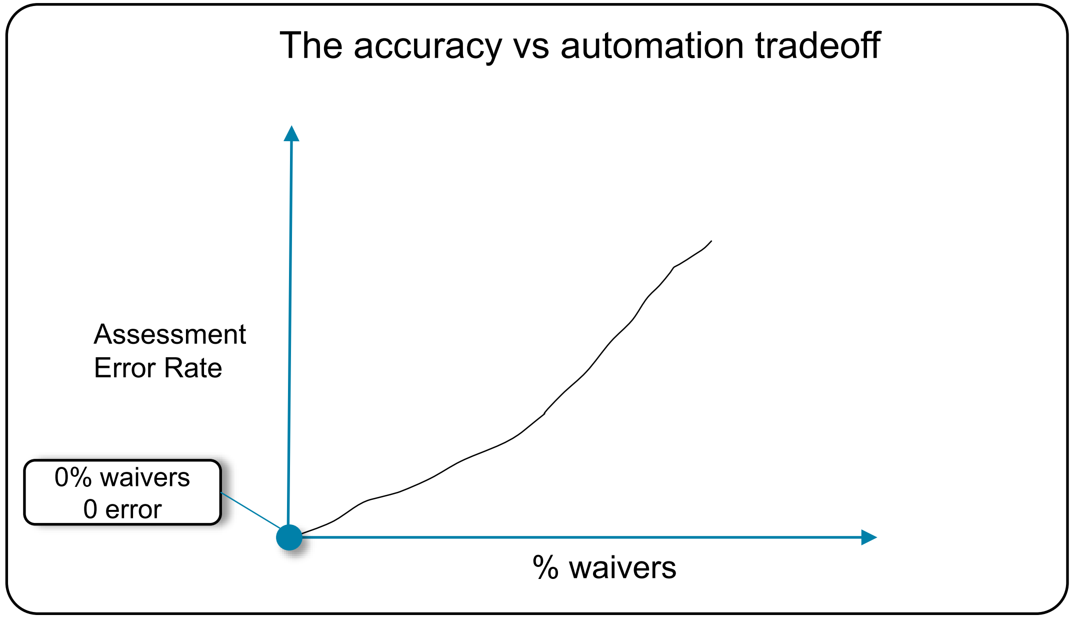


These applicants will pay less premium due to faulty risk assessment. This makes the product vulnerable to anti-selection.

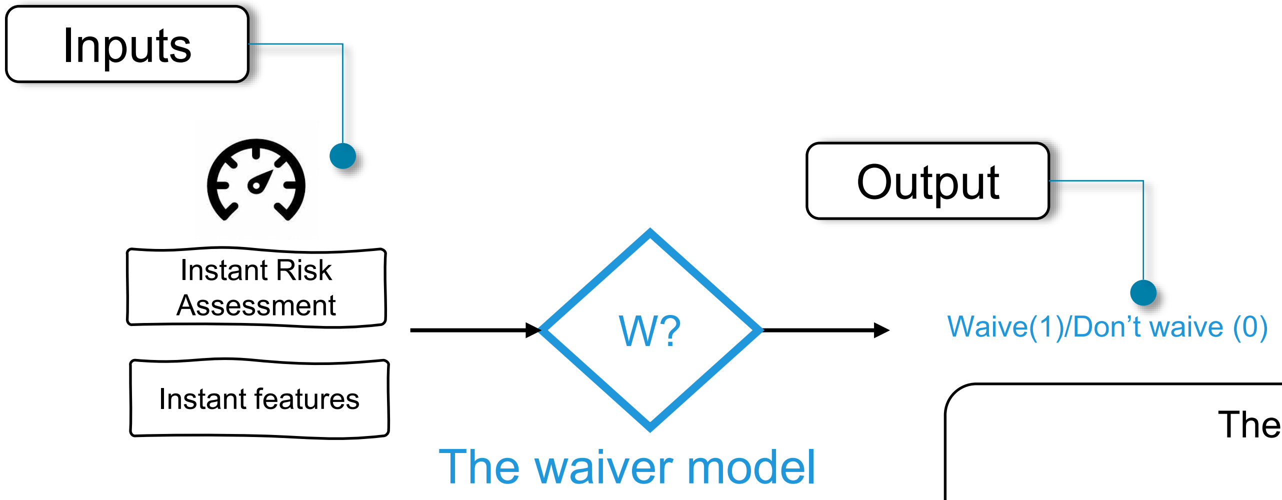
W? Risk Assessment

These applicants may feel hard done by the risk assessment they receive and will consider shopping elsewhere

Full Risk Assessment	Low	High
Low	x	
High		x



Zoom on the waiver model: performance metrics



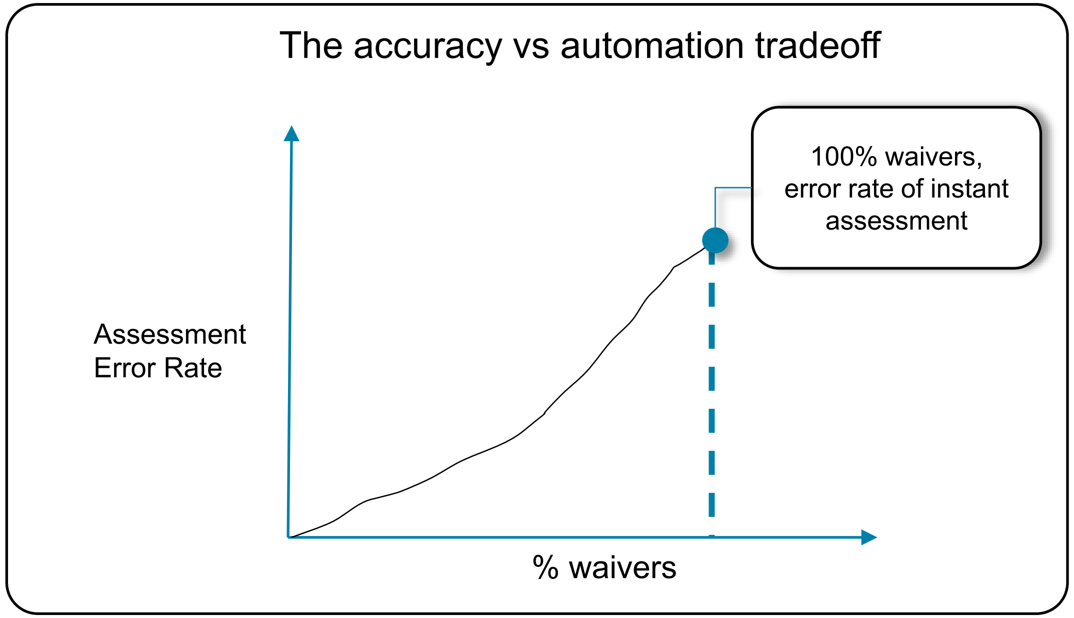
These applicants will pay less premium due to faulty risk assessment. This makes the product vulnerable to anti-selection.

W? Risk Assessment

These applicants may feel hard done by the risk assessment they receive and will consider shopping elsewhere

Full Risk Assessment

	Low	High
Low	x	
High		x



Problem Setup

Problem Setup

Multi duration (>10) historical deaths data

Training

id	age	sex	smoker	bmi	systolicbp	hba1c	triglycerides	…	exposure (years)	event
jhg76	47	M	NS	29	…	…	…	…	9.45	1
aks87	33	F	SM	24	…	…	…	…	12.1	0
…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…

100k-1M lives

Instant data features

Blood, Urine, Physical Exam features

Right Censored
Survival Labels

How to build a waiver model for new AU programs or
for replacing existing rules-based AU programs?

Problem Setup

Multi duration (>10) historical deaths data

Training — 100k-1M lives

id	age	sex	smoker	bmi	systolicbp	hba1c	triglycerides	…	exposure (years)	event
jhg76	47	M	NS	29	…	…	…	…	9.45	1
aks87	33	F	SM	24	…	…	…	…	12.1	0
…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…

Instant data features (age, sex, smoker, bmi)

Blood, Urine, Physical Exam features (systolicbp, hba1c, triglycerides, …)

Right Censored Survival Labels (exposure (years), event)

Production — 0-10k samples

Instantly data features (possibly self-disclosed, perhaps engineered differently)

(Optional) Instant Risk Assessment

Blood, Urine, Physical Exam features

id	age	sex	smoker	bmi	Instant Risk Class	systolicbp	hba1c	triglycerides	…
dff45	21	F	NS	27		…	…	…	…
nhy90	37	F	NS	24		…	…	…	…
…	…	…	…	…		…	…	…	…
…	…	…	…	…		…	…	…	…
…	…	…	…	…		…	…	…	…

Problem Setup (simulation)

Multi duration (>10) historical deaths data

Training

id	age	sex	smoker	bmi	systolicbp	hba1c	triglycerides	…	exposure (years)	event
jhg76	47	M	NS	29	…	…	…	…	9.45	1
aks87	33	F	SM	24	…	…	…	…	12.1	0
…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…

Z — Instant data features

X — Blood, Urine, Physical Exam features

Right Censored Survival Labels

100k-1M lives

Production

id	age	sex	smoker	bmi	Instant Risk Class	systolicbp	hba1c	triglycerides	…
dff45	21	F	NS	27		…	…	…	…
nhy90	37	F	NS	24		…	…	…	…
…	…	…	…	…		…	…	…	…
…	…	…	…	…		…	…	…	…
…	…	…	…	…		…	…	…	…

Z' — (Optional) Instant data features (possibly self-disclosed)

(Optional) Instant Risk Assessment

X — (Optional) Blood, Urine, Physical Exam features

0-10k samples

Problem Setup

Multi duration (>10) historical deaths data

Training

id	age	sex	smoker	bmi	systolicbp	hba1c	triglycerides	…	exposure (years)	event
jhg76	47	M	NS	29	…	…	…	…	9.45	1
aks87	33	F	SM	24	…	…	…	…	12.1	0
…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…

100k-1M lives

Instantly data features (possibly self-disclosed)

Blood, Urine, Physical Exam features

Right Censored Survival Labels

- How to build a waiver model for new AU programs?
- How to build a waiver model for replacing existing rules-based AU programs?

Production

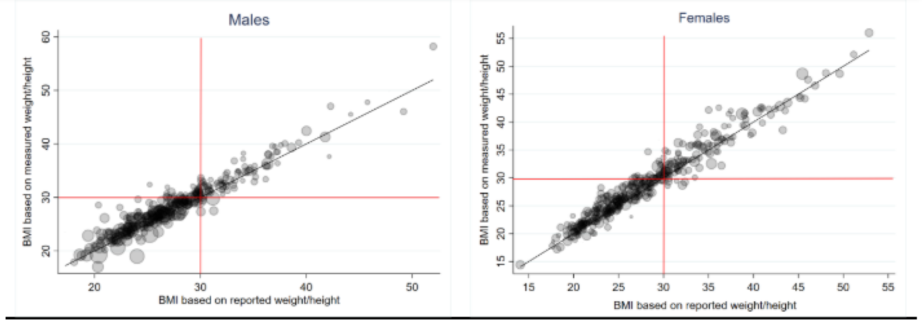
Issues:

- Limited data availability (<10k), with no mortality feedback
- No natural labels to train the waiver model (can't use historical underwriting outcomes as labels, since we don't want to regress to historical rules-based underwriting)
- Differences in instant features Z available at training time and Z' available in real world

Challenges and methods

Real world Z'

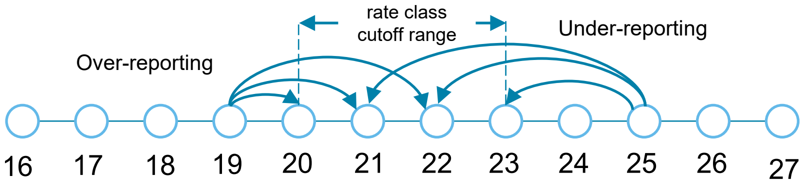
Even without any malicious intent, self-reported BMI can have errors



Notes: Markers are scaled to reflect sample weights. Darker regions representing more concentrated data points. The black line is a 45-degree line.

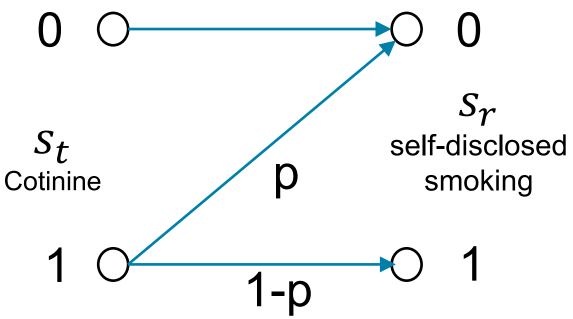
Apostolos Davillas, Andrew M. Jones, The implications of self-reported body weight and height for measurement error in BMI, Economics Letters, Volume 209, 2021, 110101, ISSN 0165-1765

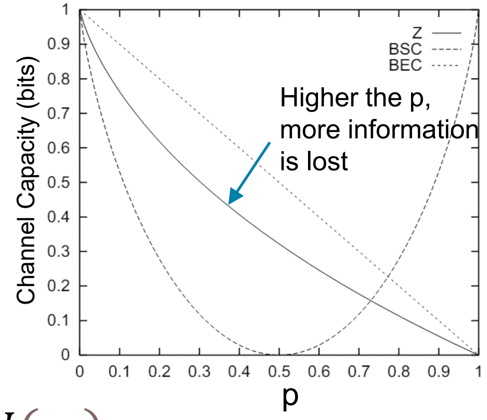
Life insurance applicants may over/under-report to get better rate class



$$\underbrace{b_r}_{\text{self-reported bmi}} = \underbrace{b_t}_{\text{measured bmi}} + \underbrace{N(b_t)}_{\text{heteroscedastic noise}} + \underbrace{\epsilon}_{\text{homoscedastic noise}}$$

Self disclosed smoking resembles cotinine presence/absence variable corrupted by the Z channel[2]





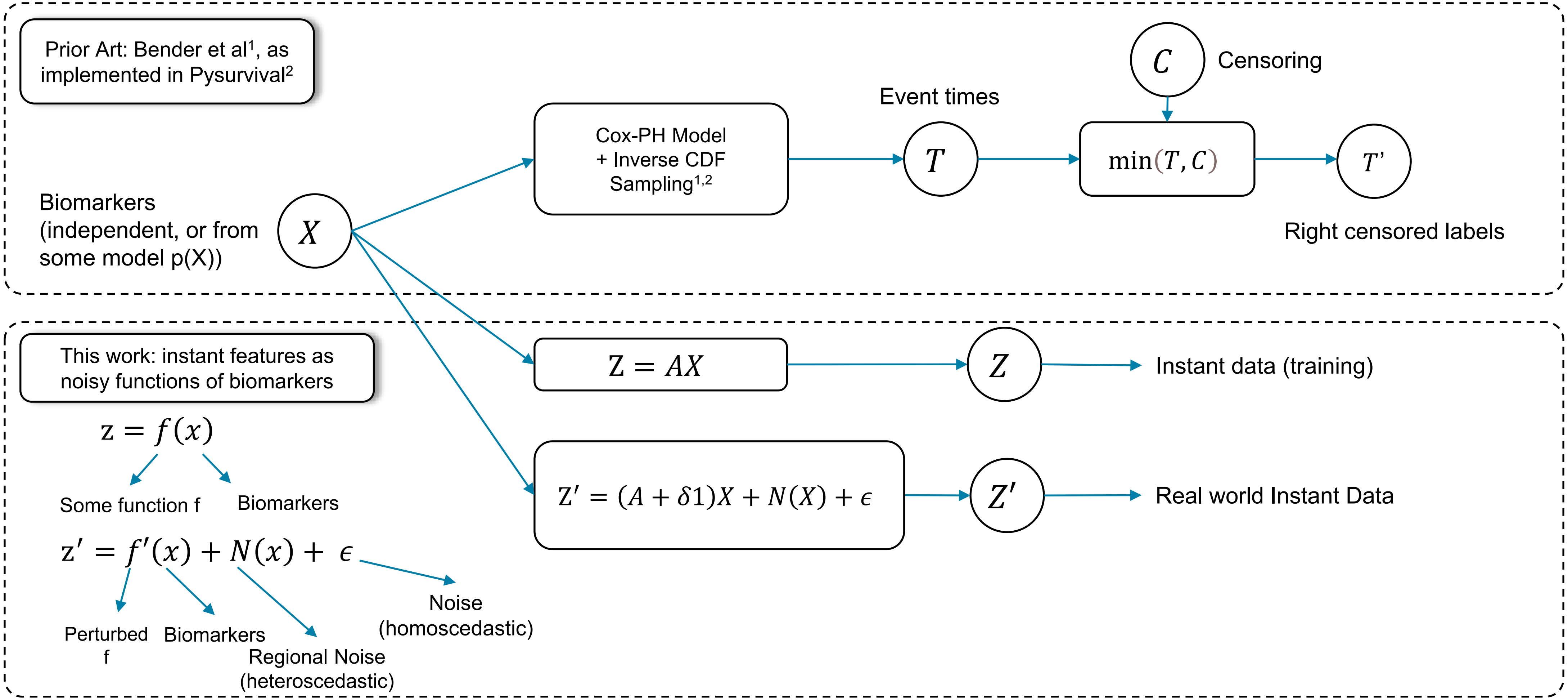
$$\underbrace{s_r}_{\text{self-reported smoking status (0/1)}} = \underbrace{s_t}_{\text{blood cotinine (0/1)}} + \underbrace{N(s_t)}_{\text{heteroscedastic noise}}$$

David J. C. MacKay. 2002. Information Theory, Inference & Learning Algorithms. Cambridge University Press, US

Application questions change often, i.e. Z' can be related to Z but slightly different

(2001) Do you smoke cigarettes on regular basis?
(2024) Do you smoke cigarettes, vape or any other tobacco product on regular basis?

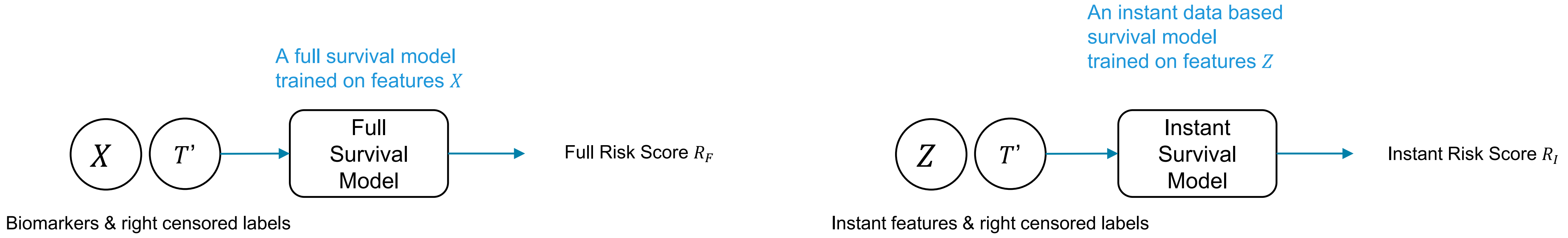
Simulation Setup: instant features as noisy functions of biomarkers



[1]Bender, R., Augustin, T., & Blettner, M. (2005). Generating survival times to simulate Cox proportional hazards models. Statistics in medicine, 24(11), 1713-1723.
[2]Fotso et al, PySurvival: Open source package for Survival Analysis modeling, 2019--, https://www.pysurvival.io/, https://github.com/square/pysurvival

Method 1: instant survival model based

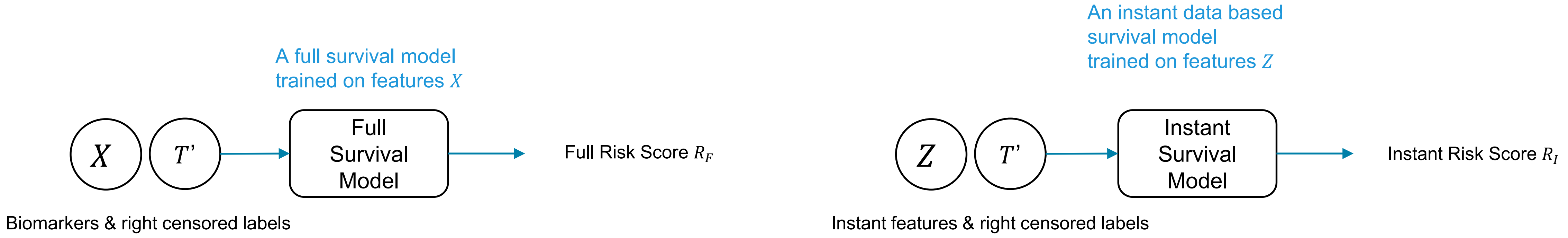
Training



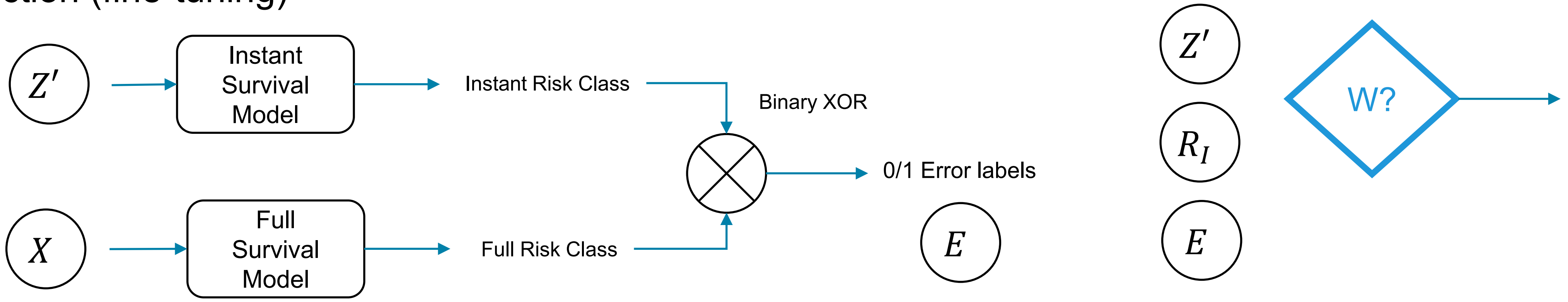
Production

Method 2: Can the full survival model help?

Training

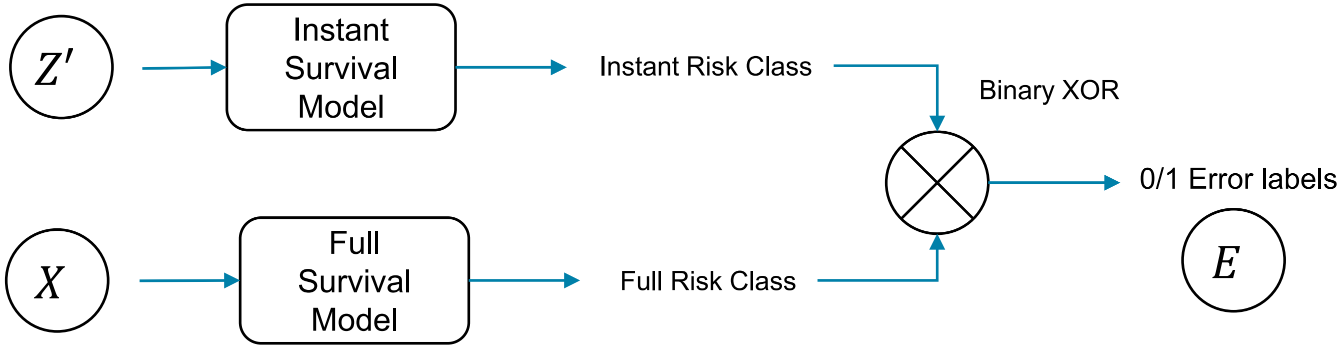


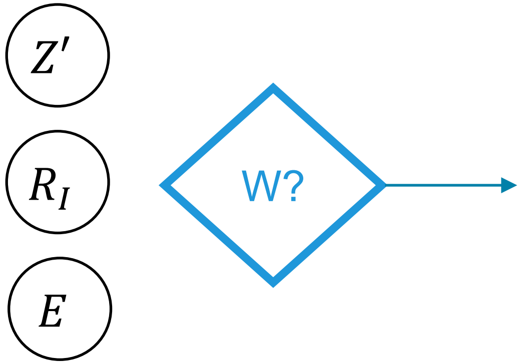
Production (fine-tuning)

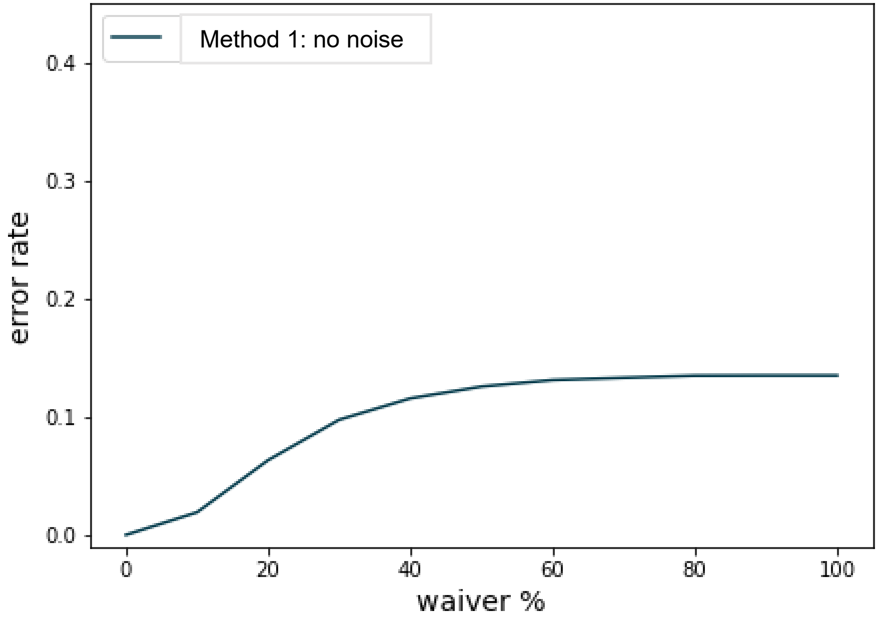


Method 2: Can the full survival model help?

Production (Fine-tuning)



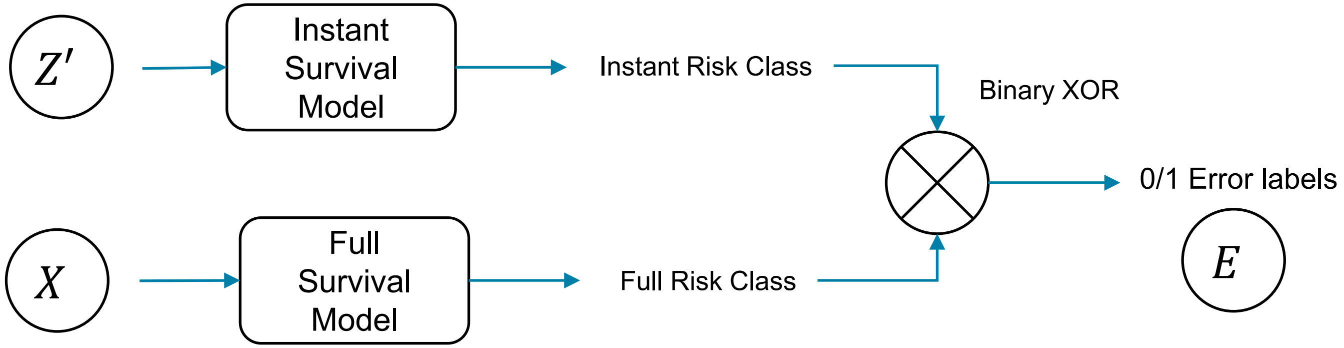


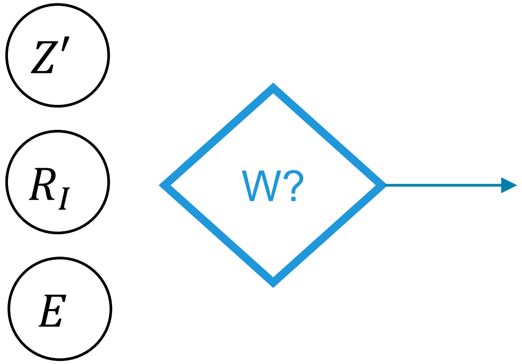


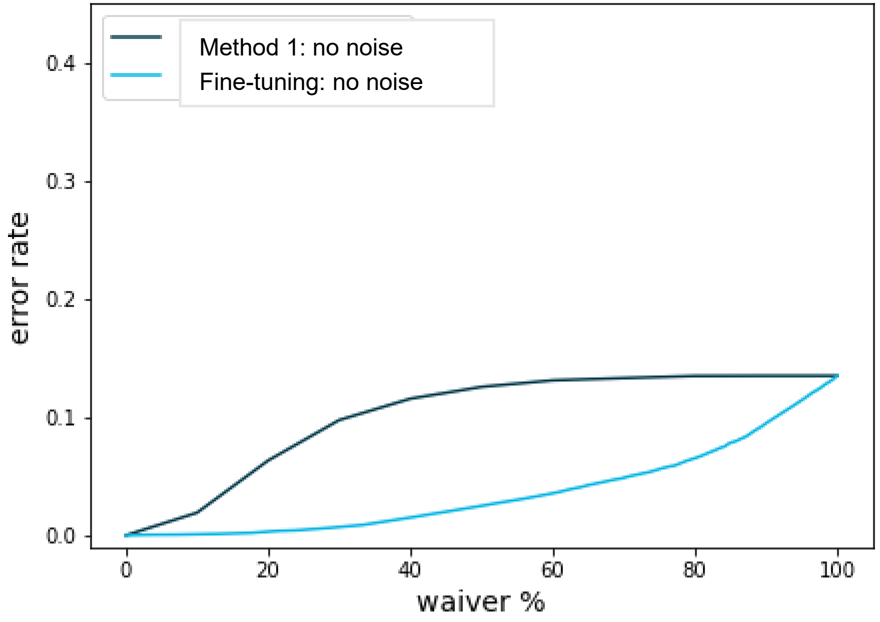
SCOR
The Art & Science of Risk

Method 2: Can the full survival model help?

Production (Fine-tuning)

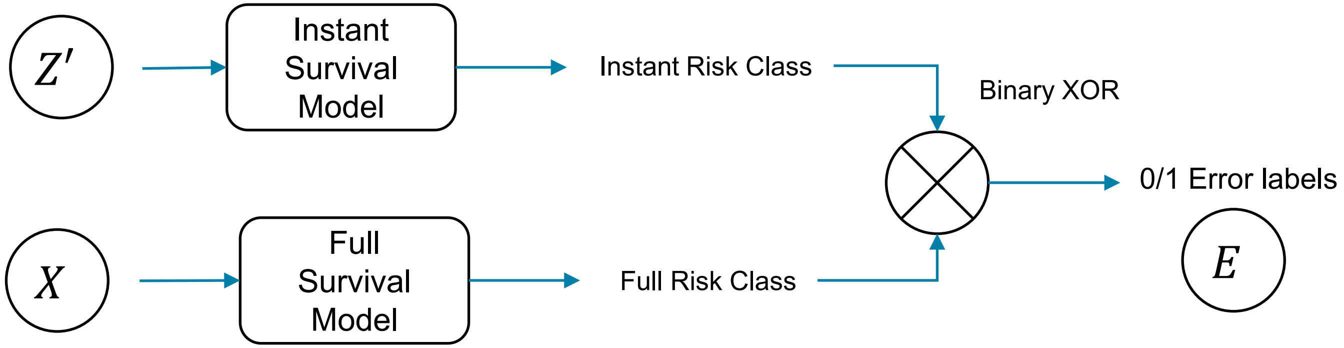


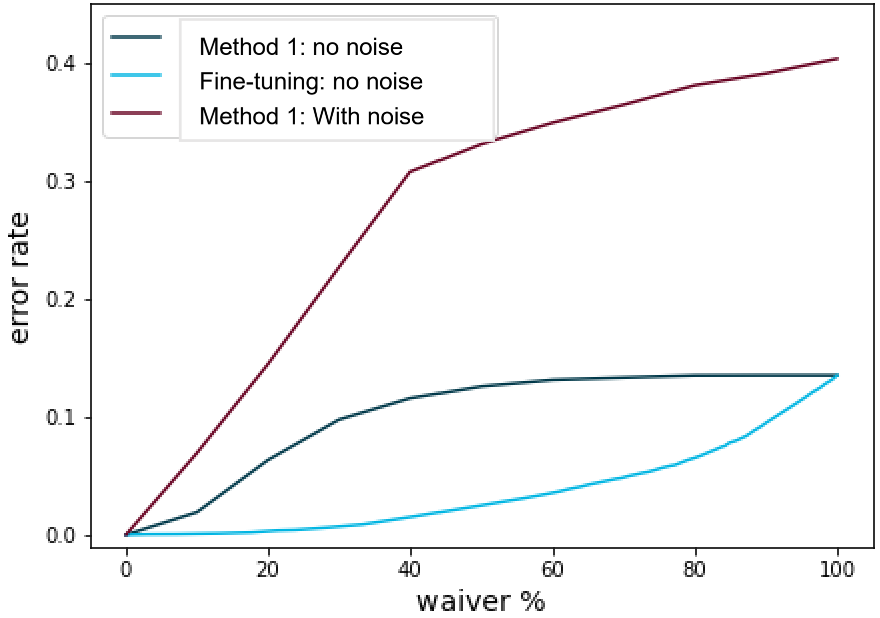


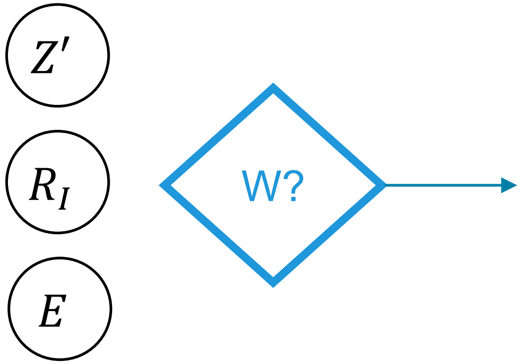


Method 2: Can the full survival model help?

Production (Fine-tuning)

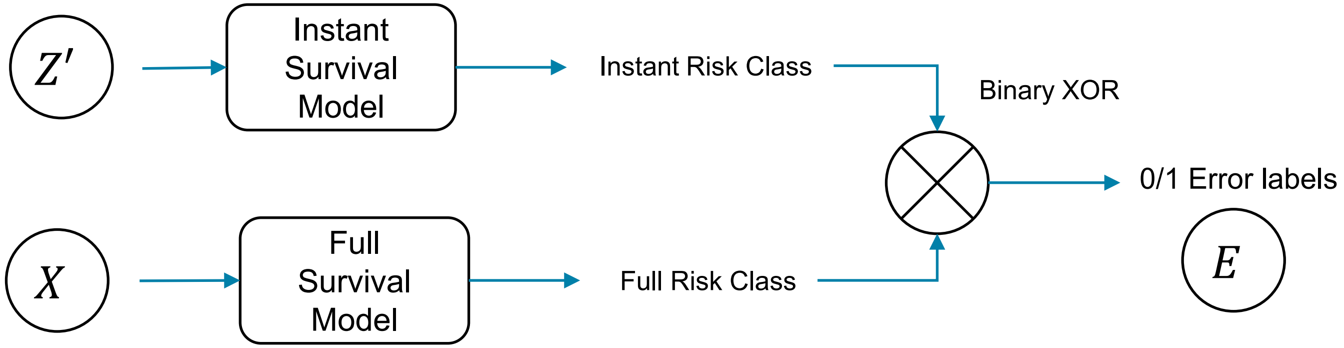


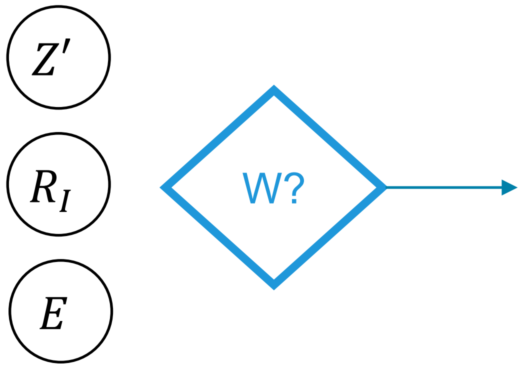


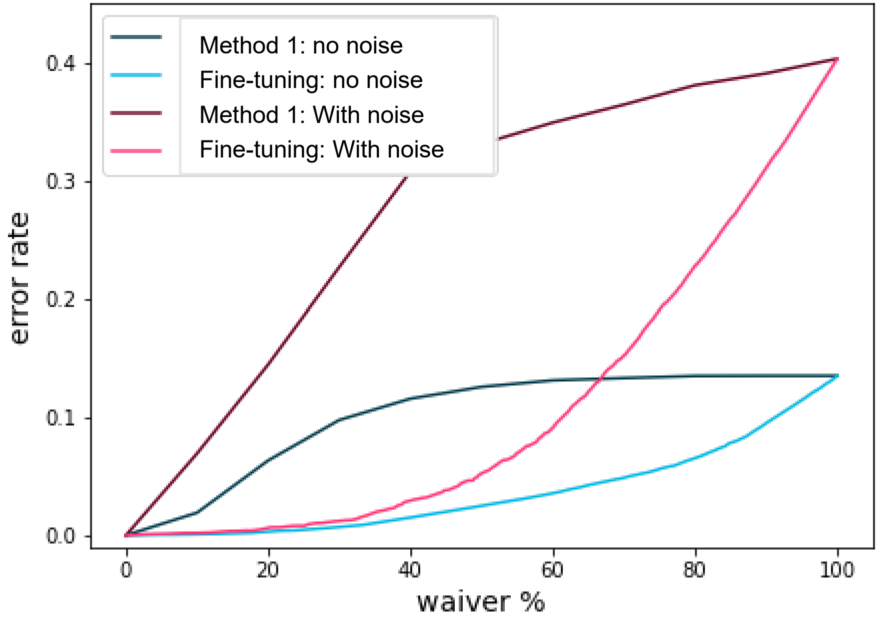


Method 2: Can the full survival model help? Yes!

Production (Fine-tuning)







What we addressed in this talk

- Intro to Algorithmic Underwriting and Accelerated Underwriting in US life market
- Intro to evidence waiver models from historical claims data
- Real world challenges with data
- Fine-tuning on a small sample of real-world data

Not addressed in this talk but important

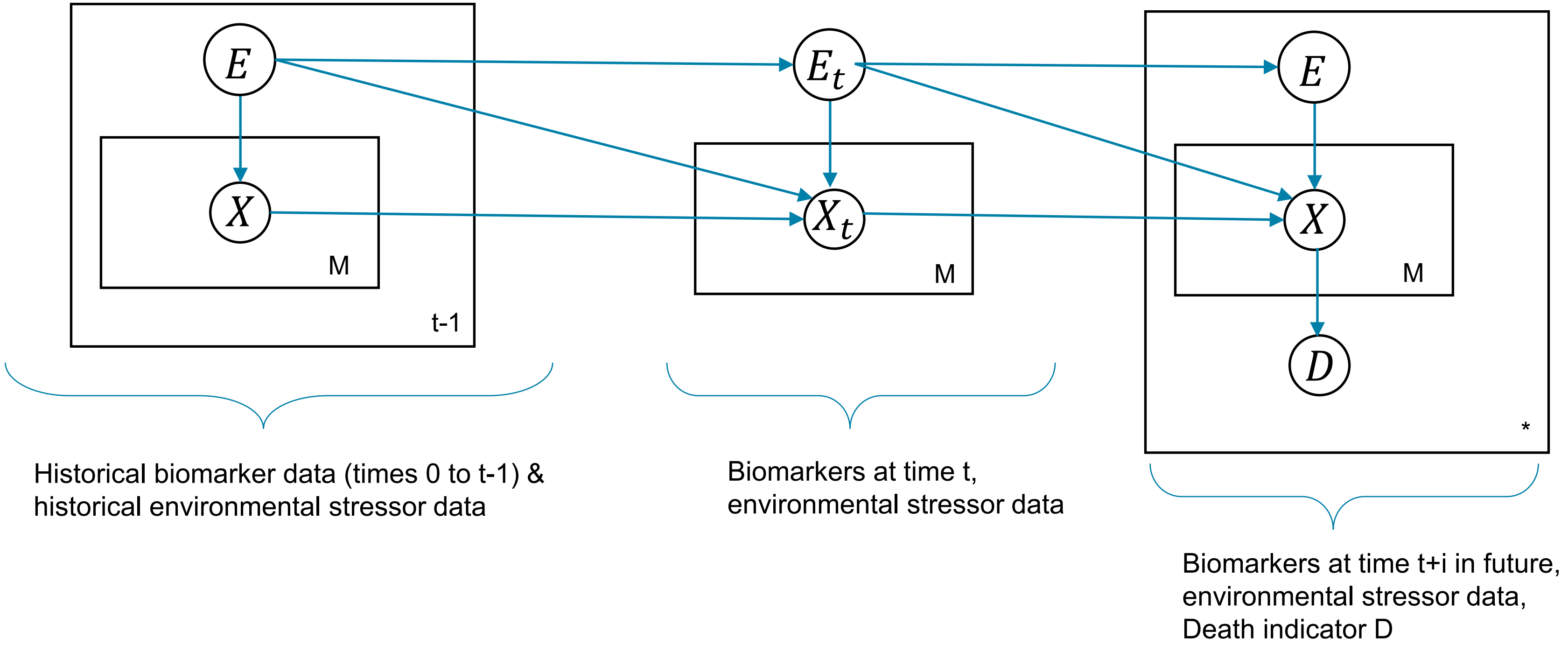
- Relationship to other concepts:
  - Knowledge Distillation
  - Learning using Privileged Information (LUPI)
  - Boosting
- Experiments with NHANES
  - Real world P(X)
  - Real world noise models for various features

SCOR The Art & Science of Risk

Thank You

Appendix

Stochastic Process Model of Mortality



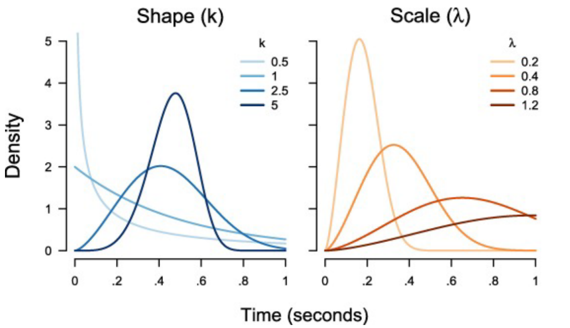
- Biomarkers
  - portmanteau of “biological marker”, refers to a broad subcategory of medical signs – that is, objective indications of medical state observed from outside the patient – which can be measured accurately and reproducibly
  - WHO: any substance, structure, or process that can be measured in the body or its products and influence or predict the incidence of outcome or disease
  - NIH: a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention
- Propopsed by Yashin et al, this is a very general model to keep in mind, as this is relevant to whether generalized distillation will help or not
- The model can be thought of as a “random walk with manholes” in the space of biomarkers and environmental stressors
- Some states labeled “manholes” (D=1)

Strimbu K, Tavel JA. What are biomarkers? Curr Opin HIV AIDS. 2010 Nov;5(6):463-6.
Yashin, A.I. *et al.* (2016). Stochastic Process Models of Mortality and Aging. In: Biodemography of Aging. The Springer Series on Demographic Methods and Population Analysis, vol 40. Springer, Dordrecht.

Generating right censored survival data

Weibull distribution with shape k>0 and scale λ>0

$$f(x) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$



k is v in our notation, while λ is a function of covariates

Formula recap

- Hazard is instantaneous danger
  $\lambda(t) = \lim_{dt \to 0} \frac{P(t \leq T \leq t + \mathrm{d}t \mid T \geq t)}{\mathrm{d}t} = \frac{f(t)}{S(t)} = -\frac{\mathrm{d}}{\mathrm{d}t} \ln[S(t)]$
- Cumulative Hazard: $M(t) = \int_0^t \lambda(u)\,\mathrm{d}u$
- PDF: distribution of the times of death: $f(t) = \lambda(t)S(t)$
- Survival function: $S(t) = 1 - F(t) = \int_t^\omega f(u)\,\mathrm{d}u = \mathrm{e}^{-M(t)}$
- Cox proportional hazards model (linear): $\lambda(t|x) = \lambda_0(t)\mathrm{e}^{\beta x}$

Case I: Constant baseline hazard function

$\lambda(t|x) = \lambda_0 \mathrm{e}^{\beta x} \rightarrow M(t) = \int_0^t \lambda_0\, \mathrm{e}^{\beta x} dt = \lambda_0 \mathrm{e}^{\beta x} t \rightarrow S(t|x) = \mathrm{e}^{-\lambda_0 \mathrm{e}^{\beta x} t} = \mathrm{e}^{-\beta'(x)t} \rightarrow f(t|x) = \beta'(x)\mathrm{e}^{-\beta'(x)t}$

i.e. the survival times $f(t|x)$ are exponentially distributed, where the rate of the exponential distribution is dependent on covariates

Case II: Weibull: Baseline hazard is a function of time specified as follows:

$\lambda(t|x) = \lambda_0 v t^{v-1} \mathrm{e}^{\beta x} \rightarrow M(t|x) = \int_0^t \lambda_0 v t^{v-1} \mathrm{e}^{\beta x} dt = \lambda_0 t^v \mathrm{e}^{\beta x} \rightarrow S(t|x) = \mathrm{e}^{-\lambda_0 t^v \mathrm{e}^{\beta x}} = \mathrm{e}^{-\beta'(x)t^v} \rightarrow f(t|x) = v t^{v-1} \beta'(x) \mathrm{e}^{-\beta'(x)t^v}$

i.e. the covariates inform the scale parameter of a Weibull distribution (~intensity of the mode) while v is the shape parameter of the distribution (~location of the mode)

Case III: Gompertz: Baseline hazard is exponential in time t

$\lambda(t|x) = \lambda_0 \mathrm{e}^{\alpha t} \mathrm{e}^{\beta x} \rightarrow M(t|x) = \int_0^t \lambda_0 \mathrm{e}^{\alpha t} \mathrm{e}^{\beta x} dt = \frac{\lambda_0 \mathrm{e}^{\alpha t} \mathrm{e}^{\beta x} - \lambda_0 \mathrm{e}^{\beta x}}{\alpha} \rightarrow S(t|x) = e^{-\left(\frac{\lambda_0 \mathrm{e}^{\alpha t} \mathrm{e}^{\beta x} - \lambda_0 \mathrm{e}^{\beta x}}{\alpha}\right)} = e^{-\frac{\beta'(x)(\mathrm{e}^{\alpha t} - 1)}{\alpha}}$

$\rightarrow f(t|x) = \beta'(x) e^{\alpha t} e^{-\frac{\beta'(x)(e^{\alpha t} - 1)}{\alpha}} = \beta'(x) e^{(\beta'(x) + \alpha)\mathrm{e}^{\alpha t} + \beta'(x)}$

CaseIV: Other distributions (see Bender et al.)

Bender, R., Augustin, T., & Blettner, M. (2005). Generating survival times to simulate Cox proportional hazards models. Statistics in medicine, 24(11), 1713-1723.

Generating right censored survival data

Cox proportional hazards model (linear):

- Formula: $\lambda(t|x) = \lambda_0(t)\mathrm{e}^{\beta x}$
- Survival function: $S(t) = \mathrm{e}^{-\int_0^t \lambda(u|x)\,\mathrm{d}u} = \mathrm{e}^{-\int_0^t \lambda_0(u)\mathrm{e}^{\beta x}\,\mathrm{d}u} = \mathrm{e}^{-\mathrm{e}^{\beta x}\int_0^t \lambda_0(u)\,\mathrm{d}u} = \mathrm{e}^{-\mathrm{e}^{\beta x}M_0(t)}$
- CDF: $F(t|x) = 1 - \mathrm{e}^{-\mathrm{e}^{\beta x}M_0(t)}$
- Data generation strategy is as follows:
  1. Generate covariates x
     a) Independently with univariate distributions of choice
     b) From some joint distribution p(x) (perhaps learnt using unsupervised learning)
  2. Assume a baseline hazard $\lambda_0(t)$ and derive $F(t|x)$
  3. Use inverse transform sampling
     - The cdf F transforms some domain into [0,1] interval i.e. $F: [a, b] \rightarrow [0,1]$ for some a,b
     - If F is invertible, we can generate a random variable with cdf F as $\mathrm{F}^{-1}(U)$ where U is uniform RV over [0,1]
     - Bender et al[1] simplify this to

$$\mathrm{T} = M_0^{-1}[-\log(U) \times e^{-\beta x}]$$

  4. Censoring time C is generated independently from $\mathcal{N}(c, 5)$, event of interest happens if T<C else it is censored
- Pysurvival[2] implements the above process for case I-III of baseline hazard and independent covariates

[1]Bender, R., Augustin, T., & Blettner, M. (2005). Generating survival times to simulate Cox proportional hazards models. Statistics in medicine, 24(11), 1713-1723.
[2]Fotso et al, PySurvival: Open source package for Survival Analysis modeling, 2019--, https://www.pysurvival.io/, https://github.com/square/pysurvival/issues/15#issuecomment-579584083

Refresher on boosting

1 F_0 is the instant survival model for our context

Input: training set $\{(x_i, y_i)\}_{i=1}^n$, a differentiable loss function $L(y, F(x))$, number of iterations M.

Algorithm:

1. Initialize model with a constant value:

$$F_0(x) = \arg\min_\gamma \sum_{i=1}^n L(y_i, \gamma).$$

2 We perform a single boosting round (M=1)

2. For m = 1 to M:
   1. Compute so-called *pseudo-residuals*:

$$r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)} \quad \text{for } i = 1, \ldots, n.$$

3

The residuals in context of fine-tuning are the binary rate class errors

   2. Fit a base learner (e.g. tree) $h_m(x)$ to pseudo-residuals, i.e. train it using the training set $\{(x_i, r_{im})\}_{i=1}^n$.
   3. Compute multiplier γ_m by solving the following one-dimensional optimization problem:

$$\gamma_m = \arg\min_\gamma \sum_{i=1}^n L\left(y_i, F_{m-1}(x_i) + \gamma h_m(x_i)\right).$$

   4. Update the model:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x).$$

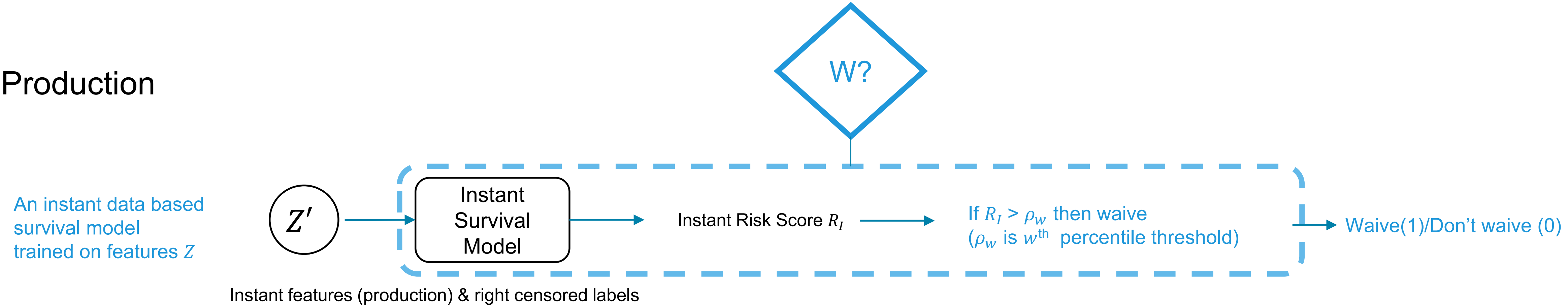
3. Output $F_M(x)$.

4 Contrary to boosting, the waiver model created using fine tuning is not additive, instead it uses survival model score as input (recursive?)

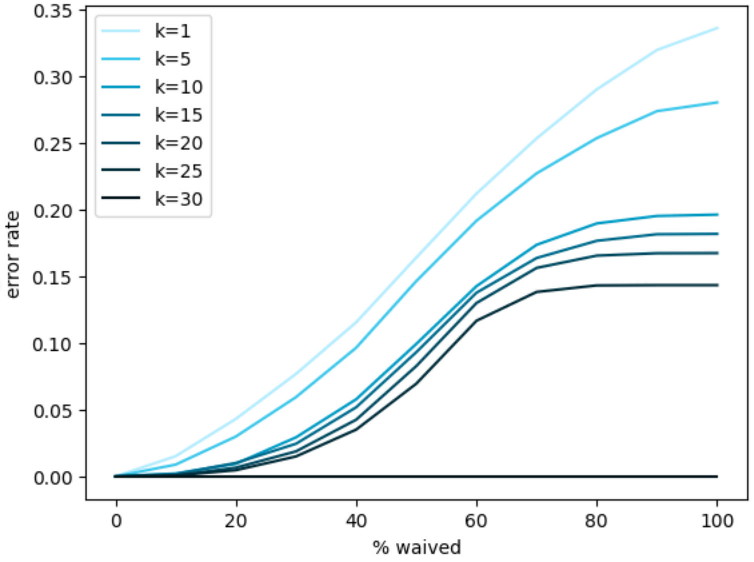
SCOR The Art & Science of Risk

Baseline: $Z' = Z$ (no noise, no concept drift)

Production

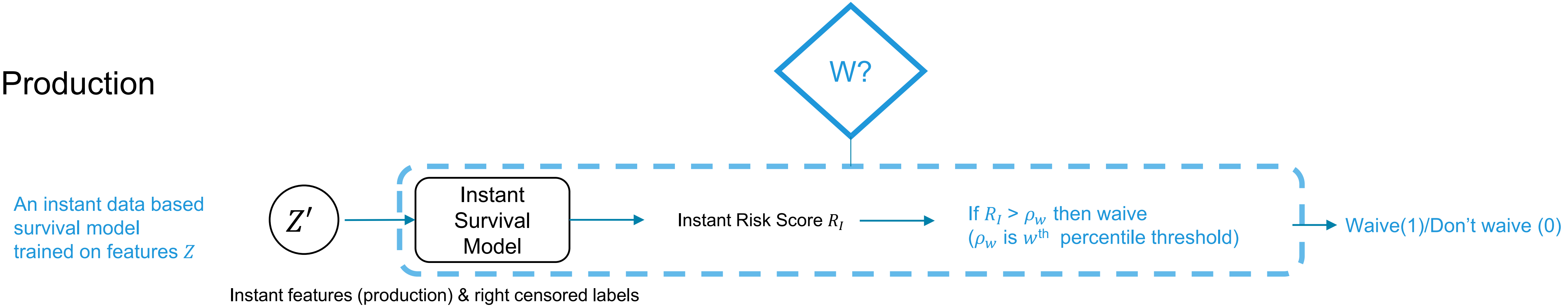


Additionally, we use ρ_{50} as the threshold to assign an instant rate class for "waived" apps and calculate error relative to full risk assessment.



Baseline: $Z' = Z$

Production



Less features ⇒ higher error