

## ARTICLE OPEN



# Assessing overdiagnosis of fecal immunological test screening for colorectal cancer with a digital twin approach

Ting-Yu Lin<sup>1,7</sup>, Sherry Yueh-Hsia Chiu<sup>2,3,7</sup>, Ling-Chun Liao<sup>4</sup>, Sam Li-Sheng Chen<sup>5</sup>, Han-Mo Chiu<sup>6,8</sup>✉ and Tony Hsiu-Hsi Chen<sup>1,8</sup>✉

Evaluating the magnitude of overdiagnosis associated with stool-based service screening for colorectal cancer (CRC) beyond a randomized controlled trial is often intractable and understudied. We aim to estimate the proportion of overdiagnosis in population-based service screening programs for CRC with the fecal immunochemical test (FIT). The natural process of overdiagnosis-embedded disease was first built up to learn transition parameters that quantify the pathway of non-progressive and progressive screen-detected cases calibrated with sensitivity, while also taking competing mortality into account. The Markov algorithms were then developed for estimating these transition parameters based on Taiwan FIT service CRC screening data on 5,417,699 residents aged 50–69 years from 2004 to 2014. Following the digital twin design with the parallel universe structure for emulating the randomized controlled trial, the screened twin, mirroring the control group without screening, was virtually recreated by the application of the above-mentioned trained parameters to predict CRC cases containing overdiagnosis. The ratio of the predicted CRCs derived from the screened twin to the observed CRCs of the control group minus 1 was imputed to measure the extent of overdiagnosis. The extent of overdiagnosis for invasive CRCs resulting from FIT screening is 4.16% (95% CI: 2.61–5.78%). The corresponding figure is increased to 9.90% (95% CI: 8.41–11.42%) for including high grade dysplasia (HGD) and further inflated to 15.83% (95% CI: 15.23–16.46%) when the removal adenoma is considered. The modest proportion of overdiagnosis modelled by the digital twin method, dispensing with the randomized controlled trial design, suggests the harm done to population-based FIT service screening is negligible.

npj Digital Medicine (2023)6:24; <https://doi.org/10.1038/s41746-023-00763-5>

## INTRODUCTION

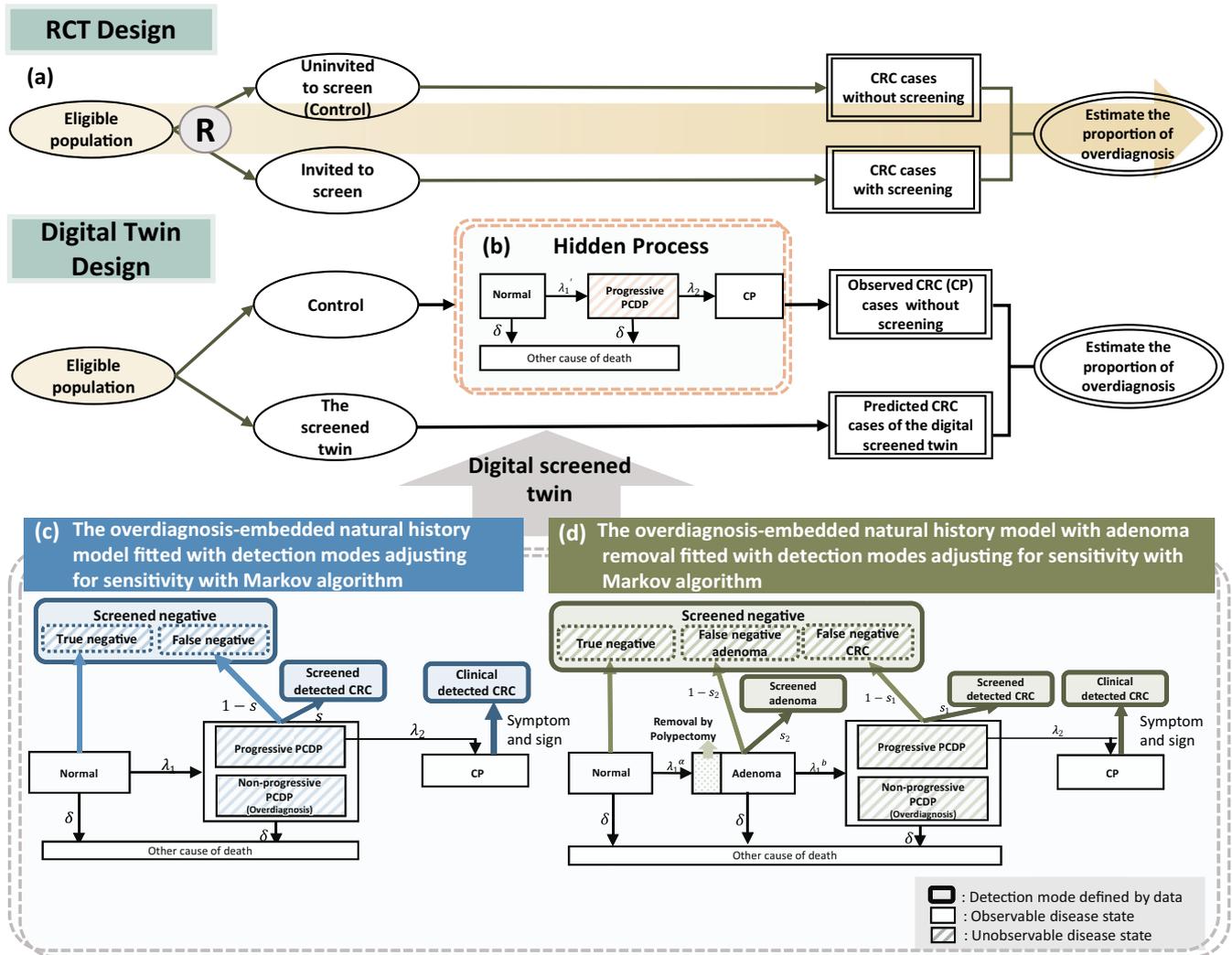
The efficacy of the guaiac fecal occult blood test (gFOBT) for colorectal cancer (CRC) screening has been proven by four randomized controlled trials (RCT) on gFOBT<sup>1–4</sup> and three service screening programs on the fecal immunological test (FIT)<sup>5–7</sup> but there still exists large uncertainty of overdiagnosis<sup>8</sup>, which is one of the demerits of population-based service screening that has already been noted in the PSA (prostate-specific antigen)<sup>9–11</sup>.

Previous studies on population-based CRC screening with gFOBT found 2%–13% overdiagnosis when the incidence method was applied to data from both the Nottingham trial and the Finnish trial<sup>12,13</sup>. Such an incidence method is valuable but must rely on the use of randomized controlled trial (RCT) data to compare the expected CRC cases derived from the screened group with those from the control group as shown in Fig. 1a. While g-FOBT has data on RCT, it still lacks RCT data for FIT screening. Evaluating overdiagnosis in two-stage FIT service screening for CRC is therefore faced with these obstacles.

The alternative to the RCT approach for assessing the extent of overdiagnosis in service screening program is to use the modelling approach as already applied to PSA screening for prostate cancer<sup>10,14,15</sup>, which takes into account the complex natural course of tumor development that is inherent to a multi-stage biological process due to environmental exposures and genetic susceptibility<sup>16</sup>, which may either accelerate or slow down the process, making allowance for competing mortality. Therefore, the definition of overdiagnosis not only comprises non-progressive cancers

that develop slowly and remain in the pre-clinical state (extremely slowly-growing tumor) but also progressive cancers that were detected in the pre-clinical state but died from other causes<sup>17</sup>. In addition to using a computer simulation approach, quantifying the extent of overdiagnosis can be escalated up to a study design level in conjunction with machine learning algorithms for estimating relevant parameters governing the overdiagnosis-embedded disease natural process. One approach relies on the recently proposed digital twin approach, which arose from the quantified-self phenomenon to promote the value of self-monitoring made possible by emerging digital technologies for data acquisition and self-reflection<sup>18</sup>. Its applications include international space station<sup>19</sup>, biology<sup>20</sup>, and medicine<sup>21</sup>. More importantly, the concept of the digital twin was realized in the realm of cancer prevention and screening by the parallel universe approach, which has already been used in a micro-simulation scenario for the development of CISNET (Cancer Intervention and Surveillance Modeling Network)<sup>22</sup>. The parallel universe structure is first to simulate the population in the absence of the screening and then simulate the identical population with the uptake of screening in order to evaluate the effectiveness of screening program<sup>23,24</sup>. The parallel universe approach, as shown in Fig. 1a, can be used to evaluate overdiagnosis by first mirroring real-world CRC cases arising from the underlying eligible population in a no-screening world and then virtually recreating the digital twin with the identical population subjected to screening. However, to build up such a digital twin requires the development of computer algorithms for

<sup>1</sup>Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan. <sup>2</sup>Department of Health Care Management, College of Management, Chang Gung University, Taoyuan, Taiwan. <sup>3</sup>Division of Hepato-Gastroenterology, Department of Internal Medicine, Kaohsiung Chang Gung Memorial Hospital, Kaohsiung, Taiwan. <sup>4</sup>Department of Pharmacy, National Taiwan University Hsin-Chu Hospital, Hsinchu, Taiwan. <sup>5</sup>School of Oral Hygiene, College of Oral Medicine, Taipei Medical University, Taipei, Taiwan. <sup>6</sup>Departments of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan. <sup>7</sup>These authors contributed equally: Ting-Yu Lin, Sherry Yueh-Hsia Chiu. <sup>8</sup>These authors jointly supervised this work: Han-Mo Chiu, Tony Hsiu-Hsi Chen. ✉email: [hanmochiu@ntu.edu.tw](mailto:hanmochiu@ntu.edu.tw); [chenlin@ntu.edu.tw](mailto:chenlin@ntu.edu.tw)



**Fig. 1 Randomized controlled trial (RCT) and digital twin design for overdiagnosis in population-based screening.** **a** The design of randomized controlled trial for evaluation of overdiagnosis; **b** the hidden disease natural history process from normal, PCDP, until the clinical phase among the control group in the absence of screening; **c** the overdiagnosis-embedded multi-state Markov model and algorithms for learning parameters with adjustment for sensitivity and competing mortality from the screened group; **d** the overdiagnosis-embedded multi-state Markov model and algorithms with adenoma for learning parameters with adjustment for sensitivity and competing mortality from the screened group.

learning parameters that govern the unobservable overdiagnosis-embedded disease natural history process.

Doing so can tackle the thorny issue of the unobservable disease natural process of distinguishing non-progressive cancers from progressive cancers that stay in the pre-clinical detectable phase (PCDP) due to the interruption of treatment. Moreover, in stool-based two-tier screening modelling for population-based service screening data on overdiagnosis through the disease natural history process would be affected by the sensitivity of the gFOBT or FIT<sup>25</sup>. The failure to calibrate the sensitivity of a test may lead to a biased estimation of overdiagnosis. Moreover, one also needs to consider the competing causes of death, which is another cause of overdiagnosis for both non-progressive and progressive cancers.

The aim of this study is therefore to first build up the overdiagnosis-embedded disease natural process governed by transition parameters and the calibrated sensitivity, while accounting for competing mortality. The corresponding Markov algorithms were developed to estimate these parameters based on the data of the population-based organized service program in Taiwan. The digital screened twin, mirroring the control group, was virtually created to predict CRC cases with the trained parameters in relation

to the overdiagnosis-embedded natural process. Those predicted cases were further compared with the real-world data on the observed CRCs for assessing the extent of overdiagnosis of CRC related to FIT screening. To have external validation of this approach, we also applied this digital twin method to gFOBT with the available data from two RCTs in the UK and Denmark, respectively.

The findings show the extent of overdiagnosis for invasive CRCs resulting from FIT screening drops from 9.90% (95% CI: 8.41–11.42%) to 4.16% (95% CI: 2.61–5.78%) with and without allowing for high-grade dysplasia (HGD), respectively. The corresponding figure is further inflated to 15.83% (95% CI: 15.23–16.46%), taking the removal of adenoma into account. With such a low proportion of overdiagnosis predicted by the digital twin method, it appears that population-based FIT service screening has suffered minimal harm.

## RESULTS

### Markov model learning process calibrated by sensitivity

Before the digital twin of 5,417,699 subjects eligible for FIT service screening in Taiwan can be virtually recreated to impute the

**Table 1.** Estimated results of overdiagnosis-embedded disease natural history model without and with consideration of adenoma removal in three studies.

Parameters	$\lambda_1$ (Normal $\rightarrow$ PCDP)	$\lambda_2$ (PCDP $\rightarrow$ Clinical)	Sensitivity <sup>c</sup>	Mean sojourn time (1/ $\lambda_2$ )
Without consideration of Adenoma removal				
Taiwan population-based study <sup>a</sup>	0.00141 (0.00139, 0.00143)	0.3860 (0.3726, 0.3997)	80.25% (78.12%, 82.32%)	2.59 (2.50, 2.68)
UK trial	0.00147 (0.00136, 0.00159)	0.3475 (0.2437, 0.4513)	53.40% (34.26%, 69.55%)	2.88 (2.22, 4.10)
Denmark trial	0.00172 (0.00155, 0.00189)	0.4433 (0.3226, 0.5639)	52.05% (35.53%, 68.56%)	2.26 (1.77, 3.10)
With consideration of Adenoma removal				
Taiwan population-based study <sup>b</sup>	$\lambda_1$ <sup>a</sup> (Normal $\rightarrow$ Adenoma)	$\lambda_1$ <sup>b</sup> (Adenoma $\rightarrow$ PCDP)	$\lambda_2$ (PCDP $\rightarrow$ CRC)	Sensitivity <sup>d</sup>
Taiwan population-based study <sup>b</sup>	0.00149 (0.00149, 0.00150)	0.0839 (0.0831, 0.0846)	0.5300 (0.5223, 0.5381)	63.48% (59.89%, 67.05%)
Taiwan population-based study <sup>b</sup>				82.86% (78.20%, 88.77%)

<sup>a</sup>Without consideration of adenoma.<sup>b</sup>With consideration of adenoma.<sup>c</sup>Sensitivity of PCDP CRC detection.<sup>d</sup>Sensitivity of Adenoma detection.

expected CRC cases for estimating the proportion of overdiagnosis, it is necessary to learn relevant parameters from the overdiagnosis-embedded disease natural history estimated by the Markov-based algorithm. We first applied the multi-state Markov model and algorithms for learning two transition parameters from this screened group shown in Fig. 1c, yielding 1.41 (per 1000 person-years) of the pre-clinical rate and 0.39 (95% CI: 0.37, 0.40) of the annual progression rate from PCDP to CP, equal to 2.59 (95% CI: 2.50, 2.68) of mean sojourn time (MST). The sensitivity was estimated as 80.25% (95% CI: 78.12%–82.32%). The lower panel of Table 1 shows the results of the corresponding estimates using the overdiagnosis-embedded disease natural history while making allowance for the influence of the removal of adenoma.

### Estimating the proportion of overdiagnosis of CRC screening

To estimate the proportion of overdiagnosis, we applied Equation (14) and Equation (15) in Supplementary Methods with the application of parameters to impute the CRC cases derived from the screened twin in comparison with the unscreened twin. For FIT-based Taiwan population-based screening, the expected CRCs were 79,469 and 84,425 without and with considering adenoma removed by polypectomy, as shown in as shown in Fig. 1c and d, which are larger than the adjusted observed 72,587 CRC cases, making allowance for the increasing incidence rate projected from the pre-screening period. When only invasive cancer was considered, the proportion of overdiagnosis as a result of FIT-based population-based screening in Taiwan was 4.16% (95% CI: 2.61–5.78%), but this increased to 9.90% (95% CI: 8.41%–11.42%) when high-grade dysplasia (HGD) was included. The overdiagnosis after further consideration of the removal of the adenoma was inflated to 15.83% (95% CI: 15.23–16.46%) (Table 2).

### Applications to two randomized controlled trials on overdiagnosis of gFOBT

Two RCT data as shown in Table 3 were used for learning parameters in relation to the imputation of the predicted cases, including progressive cancers and non-progressive cancers, as a result of the screen, as shown in Fig. 1. The upper panel of Table 2 shows that the preclinical incidence rates of RCTs with gFOBT were 1.47 and 1.72 (per 1000 person-years) for the UK and Denmark trials, respectively, and the corresponding figures of the estimated MST were 2.88 (95% CI = 2.22–4.10) and 2.26 (95% CI = 1.77–3.10) years. In addition, the estimated sensitivities of gFOBT were 53.40% (95% CI = 34.26–69.55%) for the UK trial and 52.05% (95% CI = 35.53–68.56%) for the Denmark trial.

For checking the adequacy of the proposed Markov learning model, the results of goodness of fit show that there was a lack of statistical significance with respect to the comparison between the expected and observed number of CRCs by each detection mode in two RCTs (all p-values > 0.05; Table 3). These findings suggest that the Markov model and algorithm for learning the overdiagnosis-embedded disease natural history with adjustments for sensitivity and competing mortality may be adequate to estimate the proportion of overdiagnosis.

Table 2 also shows the corresponding results of the expected CRCs in comparison with the observed CRCs. In two RCTs, the proportions of overdiagnosis due to gFOBT using the control group were 8.79% (95% CI: 8.29%–9.65%) and 9.33% (8.81%–10.20%), respectively, for the UK and Denmark trials, which are between the 2% and 13% reported in previous studies using the incidence method<sup>12,13</sup>.

### DISCUSSION

In contrast to 20% to 60% overdiagnosis<sup>26–28</sup> observed in PSA screening for prostate cancer, the proportion of overdiagnosis for FIT population-based screening is modest, implying that the harm

**Table 2.** Estimated the proportion of overdiagnosis based on overdiagnosis-embedded disease natural history model in three studies.

Study	Control group population	Mean of Follow-up time	Predicted CRC from the digital screened twin, E(t)	Observed CRC from the control group, U(t)	Overdiagnosis (%) (95% CI)
Invasive cancer only					
Taiwan population-based study <sup>a</sup>	5,417,699	11.0	66,339	63,616	4.16 (2.61,5.78) <sup>c</sup>
Including HGD					
Taiwan population-based study <sup>a</sup>	5,417,699	11.0	79,469	72,587	9.90 (8.41,11.42) <sup>c</sup>
UK trial	74,998	8.5	931	856	8.79 (8.28,9.65)
Demark trial	30,966	10.0	528	483	9.33 (8.81,10.20)
With consideration of adenoma removal					
Taiwan population-based study <sup>b</sup>	5,417,699	11.0	84,425	72,587	15.83 (15.23, 16.46) <sup>c</sup>

<sup>a</sup>Without consideration of adenoma.  
<sup>b</sup>With consideration of adenoma.  
<sup>c</sup>With consideration of competing risk of death.

**Table 3.** Number of CRC by detection modes and model validity with the overdiagnosis-embedded disease natural history model of two gFOBT randomized controlled trials of CRC screening.

Parameter	Status	Nottingham trial (UK)		Funen trial (Demark)	
		Observed	Predicted	Observed	Predicted
Screening finding by round					
Prevalent screening	Normal	44733	44735.42	20635	20630.24
	CRC	104	101.58	37	41.76
Interval cancer	CRC	164	140.04	148	147.23
	Positive but without confirmation (first round)	CRC	28	27.76	----
Positive but without confirmation (repeated round)	CRC	57	66.03	----	----
Repeated screening	Normal	88008	87977.51	66025	66014.12
	CRC	132	109.69	83	67.74
Refuser	Normal	30015	30014.06	9895	9895.71
	CRC	400	400.94	195	194.29
Internal validity Goodness of fit: Chi-square		9.9421 (d.f.=6) (p value = 0.1271)		3.9895 (d.f.=4) (p value = 0.4074)	

done to the population-based FIT service screening program is a minor seen with gFOBT. Furthermore, when only invasive cancers were included, the estimated proportion of overdiagnosis based on FIT population-based screening decreased from 9.90% to 4.16%. This finding implies that more than half of overdiagnosed CRCs were derived from HGD.

The other factor that might affect the result of overdiagnosis is the sensitivity of the test (gFOBT/FIT) which is inversely associated with the MST. Our modelling approach took sensitivity into account, and the results of the estimated sensitivities were 53.40% and 52.05% in the UK and Denmark trials, respectively. According to the previous report of the UK trial, the estimated sensitivity of gFOBT<sup>2</sup> was 53.4% by using the proportional incidence method, which was very close to our results. Our application of the overdiagnosis-embedded disease natural history model of CRC calibrated with sensitivity would render the proportion of overdiagnosis more accurate as sensitivity is also affected by the length of MST. It should be noted that the long MST is often associated with overdiagnosis if the length of follow-up time is neglected. Since sensitivity and overdiagnosis cannot be observed directly, using the overdiagnosis-embedded disease natural history model, making allowance for sensitivity, can deal with

this issue to estimate both parameters in conjunction with the pre-clinical incidence rate.

The underlying assumption of this model is that the overdiagnosed cases would not progress to CP, as shown in Fig. 1c and d mainly, including nonprogressive cancers but also progressive cancers dying from other causes of death. The sojourn time by definition is therefore infinite after the lesion entered PCDP. However, these over-detected cases derived from the control group would not be observed in the absence of screening. That accounts for why we can apply this model to estimate the proportion of overdiagnosis.

Moreover, one important characteristic of CRC screening is that the early detection of adenoma followed by its removal through polypectomy might reduce the incidence of CRC. Without considering adenoma removal, the risk of overdiagnosis may be underestimated. The advantage of using the overdiagnosis-embedded disease natural history with the incorporation of adenoma is to offset overdiagnosis contributed by the removal of adenoma by polypectomy in two-tier stool-based screening to re-estimate the incidence rate and the inflated proportion of overdiagnosis. It should be noted that the program sensitivity for adenoma may vary with different population-based screening programs<sup>29</sup>, particularly depending on the ability to detect small

or advanced adenoma by colonoscopy, which would also be reflected in subsequent screens for catching up on those missed adenomas. To assess how the calibration of sensitivity on adenoma will offset the adenoma removed by polypectomy, we did a series of sensitivity analyses by changing the parameters of sensitivity from 20% to 63%. The results show a small change from 13% to 16%. This means the consideration of adenoma removed by polypectomy with different estimates of sensitivity for detecting adenoma would only lead to a slight change in our estimate of overdiagnosis.

From the aspect of the methodology, although our developed Markov algorithm used in digital twin design as shown in Fig. 1 enables one to estimate the extent of overdiagnosis of cancer screening here it is not only the sole approach. In analogue to the digital twin approach, the parallel universe approach previously proposed for the development of CISNET is also a very useful means for evaluating the extent of overdiagnosis and would do the same task if it is applied to our data on FIT service screening<sup>23,24</sup>. The comparison of the results on different machine learning algorithms for evaluating effectiveness and overdiagnosis would enhance the verification of the digital twin design and the parallel universe structure alternative to the traditional RCT design. This would become one of the ongoing research projects in the realm of digital health.

There are several limitations to this study. First, although we consider the influence of adenoma removed by polypectomy, this does not mean we estimated the proportion of overdiagnosed adenoma, as two-tier stool-based screening can only have early detection of a proportion of adenoma rather than the whole group of adenoma. To estimate the true proportion of overdiagnosis of adenoma, it requires the same overdiagnosis-embedded model to estimate the transition parameters based on the screened group data from primary screening with colonoscopy. Otherwise, it requires long-term follow-up to remove nuisance factors such as the awareness of detecting adenoma in routine clinical practice that may affect the assessment of overdiagnosis. The second limitation is that estimating the proportion of overdiagnosis is dependent on control group information. However, in the population-based organized service screening program, there is a lack of control group without screening. Alternatively, we used the data from the screened group to project the number of cancers in comparison with the observed number using data from the pre-screening period with an adjustment for the increasing incidence rate. Whether the comparator from the pre-screening period is valid should be validated by other data on population-based service screening programs like two well-reputed ones, the Italian or Kaiser Permanente programs, both of which have demonstrated the effectiveness of the FIT screening<sup>5,6</sup>. The third limitation is that the cut-off of FIT test would determine the sensitivity, which further affects the extent of detecting non-progressive cancers (Fig. 1c and d) and the proportion of overdiagnosis. In Taiwan, the government set the threshold of 20 µg per gram of stool as the cutoff for the FIT screening program. The estimated parameters may not be directly applied to other programs with different the cut-offs. However, the proposed digital twin is still applicable if the change of sensitivity parameters in response to the alteration of cut-off can be adjusted in the overdiagnosis-embedded natural history process. The fourth concern is the psychosocial impact of overdiagnosis. Even though the proportion of overdiagnosis of CRC was estimated to be lower than 10%, there are still some concerns about anxiety and unnecessary treatment and surveillance which might cause additional economic cost and increase the risk of side effects and complications such as perforation<sup>30</sup>. However, the scope is beyond the context of this paper.

In conclusion, overdiagnosis of CRC with FIT in population-based service screening evaluated by the digital twin approach is negligible. This finding implies the FIT test is less likely to lead to

unnecessary colonoscopies and treatments when it is offered in a population-based CRC service screening program.

## METHODS

### Digital twin design

We evaluated the phenotype of overdiagnosis resulting from stool-based mass screening, with the study design borrowing from the concept of the digital twin as mentioned above. Before proposing the digital twin design, it is necessary to introduce the traditional RCT design as shown in Fig. 1a for analysis of overdiagnosis. The RCT is the standard design for evaluating whether FIT screening is effective in reducing mortality from CRC and whether there is overdiagnosis resulting from FIT screening, but it may not throw light on how overdiagnosis is invoked because the RCT neither elucidates the hidden disease natural history process from normal, through the occult pre-clinical detectable phase (PCDP) until the clinical phase in the control group in the absence of screening (Fig. 1b). Nor can the pathways leading to overdiagnosis in the screened group (Fig. 1c and d) be elucidated. Therefore, even though the result of overdiagnosis can be achieved with the RCT, the overdiagnosed cancers are often mixed up with progressive cancers detected in the pre-clinical detectable phase (PCDP) if the follow-up time is not long enough to wash out these early-detected progressive cases because of the lead time that would advance the date of diagnosis. Moreover, sensitivity is the other factor affecting the estimate of overdiagnosis when it comes to population-based service screening rather than the RCT program because the sensitivity of FIT would affect the estimate of overdiagnosis, but such an influence can't be directly observed.

To assess the extent of overdiagnosis in population-based service screening, we need to consider the separation of lead-time gained progressive cancers from non-progressive cancers and the sensitivity of FIT. The first initiative is to build up the overdiagnosis-embedded disease natural history model of CRC to distinguish progressive cancers from non-progressive ones staying in pre-clinical detectable phase (PCDP) (Fig. 1c) and then to model the unobservable process from PCDP to clinical phase (CP) for progressive cancers due to the interruption of treatment and also the removal of adenoma by polypectomy (Fig. 1d). Allowance would also be made for the sensitivity of FIT in Fig. 1c and d in order to fit different detection modes defined by the data. Relevant transition parameters include the incidence rate of PCDP (including progressive and non-progressive phenotypes), the progression rate from progressive PCDP to CP without or with considering the removal of adenoma by polypectomy, as shown in Fig. 1c and d, and sensitivity. These parameters were then trained by fitting the developed Markov machining algorithm to population-based FIT service screening data with different detection modes, as information on overdiagnosis can only be learned from those with the uptake of screening for detecting both progressive and non-progressive CRC in the PCDP as shown in Fig. 1c and d. Note that Markov algorithms developed for learning these transition parameters are detailed in the statistical section.

The spirit of the digital twin design is to emulate the RCT design to yield CRCs in the invited group and the control group through the pathways of multistate disease's natural process, as opposed to the conventional RCT design without the detailed underlying disease process. To create the mirror of the control group without screening, these trained transition parameters were further applied to generating a virtually-created digital screened twin for predicting CRCs in the identical control group subjected to screening (Fig. 1c and d). The observed CRC cases, if there are symptoms and signs related to CRC, from the control group would end up being diagnosed as CRC in the CP following the hidden

natural course of Fig. 1b from normal through the occult PCDP until the CP. Without screening, PCDP cannot be detected, nor can overdiagnosed CRC be detected in the control group. The comparison was therefore made between the predicted CRCs of the digitally screened twin and the observed CRCs of the control group in the absence of screening.

### The underlying mechanism for quantifying the proportion of overdiagnosis

The incidence rate of PCDP in Fig. 1c (denoted by  $\lambda_1$ ) is thought to be higher than in Fig. 1b (denoted by  $\lambda_1'$ ) due to an excess of nonprogressive PCDP cancers detected by screen. Cancers predicted on the basis of the virtual, digitally screened twin after applying the transition parameters learned from Fig. 1c are therefore expected to be larger than their counterparts in the control group following the natural history of Fig. 1b in the absence of screening. The excess of proportion, as calculated in the statistical section, reflects the extent of overdiagnosis.

Overdiagnosis would be underestimated if the removal of adenoma was not considered because screening not only interrupts the natural history from PCDP to CP but also stops the malignant transformation from adenoma to cancer via polypectomy. In order to adjust for such an influence, Fig. 1d shows how to apply the multi-state model, namely Normal  $\rightarrow$  Adenoma  $\rightarrow$  PCDP  $\rightarrow$  CP, to estimating three parameters in order to offset overdiagnosis contributed by the removal of adenoma by polypectomy. The predicted of CRCs of the digital screened twin based on three transition parameters ( $\lambda_1^a$ ,  $\lambda_1^b$ , and  $\lambda_2$ ) in Fig. 1d were compared with the observed CRCs of the control group in the absence of screening following the hidden natural history of Fig. 1b to calculate the proportion of overdiagnosis adjusting for the effect of removal of adenoma. Note that competing mortality, one of the causes related to overdiagnosis, was also considered for both progressive and non-progressive stays in the PCDP.

### Data source

**Taiwan Population-based CRC Screening Program with Fecal Immunochemical Test (FIT).** Population-based CRC screening data were collected from the Taiwan Colorectal Cancer Screening Program, which was launched by the Taiwanese government in 2004 and provided a biennial FIT screening for residents aged 50–69 (the upper age limit was extended to 74 after 2013). Positive cases were referred for diagnostic colonoscopy if the hemoglobin concentration was greater than 20  $\mu\text{g}$  per gram of stool (equivalent to 100 ng/mL for the Eiken OC-Sensor and 8 ng/mL for the Kyowa HM-JACK)<sup>31</sup>. For those subjects with negative result, they were advised to participate in the subsequent round of screening two years later. All screening data and the examination findings including adenoma were collected in the centralized database of the organized service screening program under the auspices of the health authority. Information on interval CRCs and CRCs from FIT screening non-participants were obtained from the National Death Registry and the Taiwan Cancer Registry according to International Classification of Diseases (ICD), 9th revision. The population cohort recruited 5,417,699 eligible population aged 50 to 69 years during the period of 2004–2009 and there were 3,811,011 participants in this program. The cohort was followed up on until 2014 to determine cancer status via a link to Taiwan's national cancer registry. A total of 71,543 CRCs were identified, including 64,199 invasive CRCs and 7,344 HGD, which were used for assessing the proportion of overdiagnosis for invasive CRCs only and all CRCs. By detection modes, there were 13,821 screen-detected CRCs, 11,904 interval cancers, and 45,818 CRCs among non-participants. The cohort was divided into two scenarios based on the overdiagnosis-embedded natural history process without considering the removal of adenoma following Fig. 1c and the overdiagnosis-embedded natural history process

with the consideration of the removal of adenoma following Fig. 1d. Because population-based service screening does not have a control arm as designed in the RCT, we used the pre-screen period data (approximate 5,417,699 population size) between 1998 and 2003 before nationwide service screening as the comparator, with an adjustment for the increasing incidence trend of CRC that would capture the biological growth rate of CRC incidence (4.5% per year) in the absence of screening<sup>7</sup>. This increasing trend was further applied to imputing the observed CRCs of this control group without screening in parallel with the contemporaneous period of the screened group serving between 2004 and 2014, as mentioned above. This study was approved by the Health Promotion Administration of the Ministry of Health and Welfare of the Taiwanese government, and informed consent was waived for the deidentified data.

### Statistical analysis

We proposed a Markov process to model the overdiagnosis-embedded natural history process in Figures c and d. Here, we learned the corresponding parameters based on population-based FIT screening data in Taiwan that consisted of both progressive and non-progressive PCDP cancers. Because the screening test (i.g. gFOBT or FIT) may miss the lesion, we took the sensitivity into account to adjust the annual progression rate ( $\lambda_2$ )<sup>32,33</sup>. As the competing risk for other causes of death ( $\delta$ ) is also one of the causes accounting for overdiagnosis, this state was incorporated into the overdiagnosis-embedded natural model in Figures c and d. Supplementary Methods give the details of how to estimate the relevant parameters with the Markov-based algorithms, namely the likelihood functions, of the multistate model after learning from the empirical data on population-based service screening by detection modes<sup>34</sup>. The Bayesian MCMC algorithm with inverse-gamma (0.001, 0.001) prior distributions for the transition parameters is used for estimating the parameters for multi-state Markov model, and the 95% credible interval is further obtained from the posterior distribution of each parameter.

As mentioned earlier, due to the overdiagnosis of CRCs, it is postulated that the annual incidence rate of CRC ( $\lambda_1$ ) applied to the digitally screened twin would be greater than that ( $\lambda_1'$ ) following the hidden natural history process arising from the control group in the absence of screening. CRC cases derived from the screened twin ( $S$ ) were virtually imputed by the already learned transition probabilities of different types of detection modes, as mentioned above, with the identical follow-up time of the control group. Therefore, we then imputed the ratio of the predicted CRCs of the screened twin ( $S(t)$ ) to the observed ( $U(t)$ ) CRCs of the control group during time  $t$ , which can be expressed by  $S(t)/U(t)$ . Then, the proportion of overdiagnosis of CRC given time  $t$  can be calculated as  $(S(t)/U(t)-1)\times 100\%$ , representing the extra proportion of screening-detected cases that would not have been diagnosed in the absence of screening. Furthermore, with the influence of the adenoma removed by polypectomy in two-tier stool-based screening, the predicted CRCs ( $S$ ) are recalculated by applying the transition parameters as shown in Fig. 1d. The detailed procedure for quantifying the proportion of overdiagnosis refers to Supplementary Methods.

We used two gFOBT randomized controlled trials to examine whether the predicted numbers of CRC based on the proposed Markov algorithm are in good agreement with the observed numbers of CRC by the detection modes using the goodness of fit of Pearson Chi-square test to validate whether the Markov-based algorithms for virtually re-creating the digital twin are adequate. The estimated proportion of overdiagnosis using the proposed digital twin method was also calculated for the gFOBT test in comparison with those estimates based on the randomized controlled trial data. The details of aggregate data by detection mode are listed in Table 3. An illustration of calculating the

proportion of overdiagnosis using the UK trial on gFOBt is also specified in Supplementary Methods.

## Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## DATA AVAILABILITY

Aggregated data that supports the findings of this study may be available upon request from the corresponding author [H.H.C.].

Individual data supporting the study's findings are not publicly available due to participant privacy.

## CODE AVAILABILITY

All statistical analyses were conducted using SAS software (Version 9.4; SAS Institute Inc. 2012). The code will be shared upon reasonable request by the first author [T.Y.L.].

Received: 27 May 2022; Accepted: 21 January 2023;

Published online: 10 February 2023

## REFERENCES

- Kronborg, O., Fenger, C., Olsen, J., Jørgensen, O. D. & Søndergaard, O. Randomised study of screening for colorectal cancer with faecal-occult-blood test. *Lancet* **348**, 1467–1471 (1996).
- Hardcastle, J. D. et al. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. *Lancet* **348**, 1472–1477 (1996).
- Mandel, J. S. et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. *N. Engl. J. Med.* **328**, 1365–1371 (1993).
- Lindholm, E., Brevinge, H. & Haglund, E. Survival benefit in a randomized clinical trial of faecal occult blood screening for colorectal cancer. *J. Br. Surg.* **95**, 1029–1036 (2008).
- Zorzi, M. et al. Impact on colorectal cancer mortality of screening programmes based on the faecal immunochemical test. *Gut* **64**, 784–790 (2015).
- Levin, T. R. et al. Effects of organized colorectal cancer screening on cancer incidence and mortality in a large community-based population. *Gastroenterology* **155**, 1383–1391 (2018).
- Chiu, H. M. et al. Long-term effectiveness of faecal immunochemical test screening for proximal and distal colorectal cancers. *Gut* **70**, 2321–2329 (2021).
- Kalager, M. et al. Overdiagnosis in colorectal cancer screening: time to acknowledge a blind spot. *Gastroenterology* **155**, 592–595 (2018).
- Pashayan, N., Powles, J., Brown, C. & Duffy, S. W. Excess cases of prostate cancer and estimated overdiagnosis associated with PSA testing in East Anglia. *Br. J. Cancer* **95**, 401–405 (2006).
- Telesca, D., Etzioni, R. & Gulati, R. Estimating lead time and overdiagnosis associated with PSA screening from prostate cancer incidence trends. *Biometrics* **64**, 10–19 (2008).
- US Preventive Services Task Force. Screening for prostate cancer: US Preventive Services Task Force recommendation statement. *Jama* **319**, 1901–1913 (2018).
- Wieszycy, P. et al. Estimation of overdiagnosis in colorectal cancer screening with sigmoidoscopy and faecal occult blood testing: comparison of simulation models. *BMJ open* **11**, e042158 (2021).
- Hakama, M. & Malila, N. Estimation of overdiagnosis by detection to incidence ratio in cancer screening. *J. Med. Screen.* **26**, 135–138 (2019).
- Davidov, O. & Zelen, M. Overdiagnosis in early detection programs. *Biostatistics* **5**, 603–613 (2004).
- Wu, G. H. M. et al. A stochastic model for survival of early prostate cancer with adjustments for leadtime, length bias, and over-detection. *Biom. J.* **54**, 20–44 (2012).
- Ushijima, T., Clark, S. J. & Tan, P. Mapping genomic and epigenomic evolution in cancer ecosystems. *Science* **373**, 1474–1479 (2021).
- Rivers, J. K. & Wulkan, S. The case for early detection of melanoma. *J. Cutan. Med. Surg.* **14**, 24–29 (2010).
- Schwartz, S. M., Wildenhaus, K., Bucher, A. & Byrd, B. Digital twins and the emerging science of self: Implications for digital health experience design and “small” data. *Frontiers in Computer. Science* **2**, 31 (2020).
- Löblich, M. & Jeggo, P. A. Hazards of human spaceflight. *Science* **364**, 127–128 (2019).

- Laubenbacher, R., Sluka, J. P. & Glazier, J. A. Using digital twins in viral infection. *Science* **371**, 1105–1106 (2021).
- Kovatchev, B. A century of diabetes technology: signals, models, and artificial pancreas control. *Trends Endocrinol. Metab.* **30**, 432–444 (2019).
- Van Ravesteyn, N. T. et al. Benefits and harms of mammography screening after age 74 years: model estimates of overdiagnosis. *J. Natl Cancer Inst.* **107**, djv103 (2015).
- Lansdorp-Vogelaar, I. et al. Personalizing age of cancer screening cessation based on comorbid conditions: model estimates of harms and benefits. *Ann. Intern. Med.* **161**, 104–112 (2014).
- van den Broek, J. J., van Ravesteyn, N. T., Heijnsdijk, E. A. & de Koning, H. J. Simulating the Impact of Risk-Based Screening and Treatment on Breast Cancer Outcomes with MISCAN-Fadia. *Med. Decis. Mak.: Int. J. Soc. Med. Decis. Mak.* **38**, 545–655 (2018).
- Elsafi, S. H., Alqahtani, N. I., Zakary, N. Y. & Al Zahrani, E. M. The sensitivity, specificity, predictive values, and likelihood ratios of fecal occult blood test for the detection of colorectal cancer in hospital settings. *Clin. Exp. Gastroenterol.* **8**, 279 (2015).
- Draisma, G. et al. Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. *J. Natl Cancer Inst.* **101**, 374–383 (2009).
- Pathirana, T., Hayden, A., Doust, J., Glasziou, P. & Bell, K. Lifetime risk of prostate cancer overdiagnosis in Australia: quantifying the risk of overdiagnosis associated with prostate cancer screening in Australia using a novel lifetime risk approach. *BMJ open* **9**, e022457 (2019).
- Yen, A. M. F. & Chen, H. H. Modeling the overdiagnosis of screen-identified cancers in population-based cancer screening with the Coxian phase-type Markov process. *Stat. Med.* **39**, 660–673 (2020).
- van der Meulen, M. P., Lansdorp-Vogelaar, I., van Heijningen, E. M., Kuipers, E. J. & van Ballegooijen, M. Nonbleeding adenomas: Evidence of systematic false-negative fecal immunochemical test results and their implications for screening effectiveness-A modeling study. *Cancer* **122**, 1680–1688 (2016).
- Toft, E. L., Kaae, S. E., Malmqvist, J. & Brodersen, J. Psychosocial consequences of receiving false-positive colorectal cancer screening results: a qualitative study. *Scand. J. Prim. Health Care* **37**, 145–154 (2019).
- Booth, D. M. et al. Reactive oxygen species induced by bile acid induce apoptosis and protect against necrosis in pancreatic acinar cells. *Gastroenterology* **140**, 2116–2125 (2011).
- Chen, T. H. H. et al. Evaluation of a selective screening for colorectal carcinoma: the Taiwan Multicenter Cancer Screening (TAMCAS) project. *Cancer: Interdiscip. Int. J. Am. Cancer. Soc.* **86**, 1116–1128 (1999).
- Chiu, S. Y. H. et al. Effect of baseline breast density on breast cancer incidence, stage, mortality, and screening parameters: 25-year follow-up of a Swedish mammographic screening. *Cancer Epidemiol. Prev. Biomark.* **19**, 1219–1228 (2010).
- Chen, T. H. H. et al. Estimation of sojourn time in chronic disease screening without data on interval cases. *Biometrics* **56**, 167–172 (2000).

## ACKNOWLEDGEMENTS

This work was supported by the National Science and Technology Council, Taiwan (grant number 111-2118-M-002-004 -MY2, 111-2811-M-002-061 -MY2, NSC 102-2314-B-182-028 -MY3), and Chang Gung Memorial Hospital and Chang Gung University (grant number BMRP-B81). The National Science and Technology Council of Taiwan (grant numbers 111-2118-M-002-004 - MY2, 111-2811-M-002-061 - MY2, NSC 102-2314-B-182-028 - MY3), Chang Gung Memorial Hospital, and Chang Gung University (grant number BMRP-B81) supported this research.

## AUTHOR CONTRIBUTIONS

H.H.C. and H.M.C. contributed to applying the digital twin approach to design to population-based FIT service screening, T.Y.L. and Y.H.C. contributed to building up the multistate overdiagnosed natural history model and developing the Markov machine learning algorithm. T.Y.L., Y.H.C., L.C.L. and L.S.C. conducted statistical analysis of the data for estimating the parameters. T.Y.L. and Y.H.C. first drafted the manuscript. H.M.C. and H.H.C. contributed to revising the manuscript. All authors made interpretations of the data and results and approved the final version of the manuscript. T.Y.L. and Y.H.C. contributed equally to this work. H.M.C. and H.H.C. are jointly supervising the study.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains Supplementary Methods available at <https://doi.org/10.1038/s41746-023-00763-5>.

**Correspondence** and requests for materials should be addressed to Han-Mo Chiu or Tony Hsiu-Hsi Chen.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023