Metamorphic Testing and Debugging of Tax Preparation Software

Saeid Tizpaz-Niari
Computer Science Department
University of Texas at El Paso

Verya Monjezi
Computer Science Department
University of Texas at El Paso

Morgan Wagner
Psychology Department
University of Texas at El Paso

Shiva Darian
Information Science Department
University of Colorado Boulder

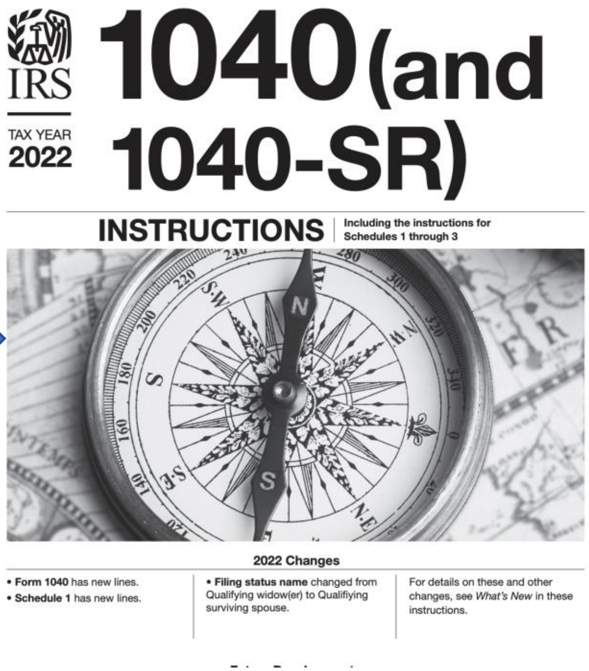
Krystia Reed
Psychology Department
University of Texas at El Paso

Ashutosh Trivedi
Computer Science Department
University of Colorado Boulder

“ Our new Constitution is now established, and has an appearance that promises permanency; but in this world nothing can be said to be certain, except death and taxes.”

— Benjamin Franklin, in a letter to Jean-Baptiste Le Roy, 1789

U.S. Tax 101: Manual Tax Filling



IRS

TAX YEAR 2022

1040 (and 1040-SR)

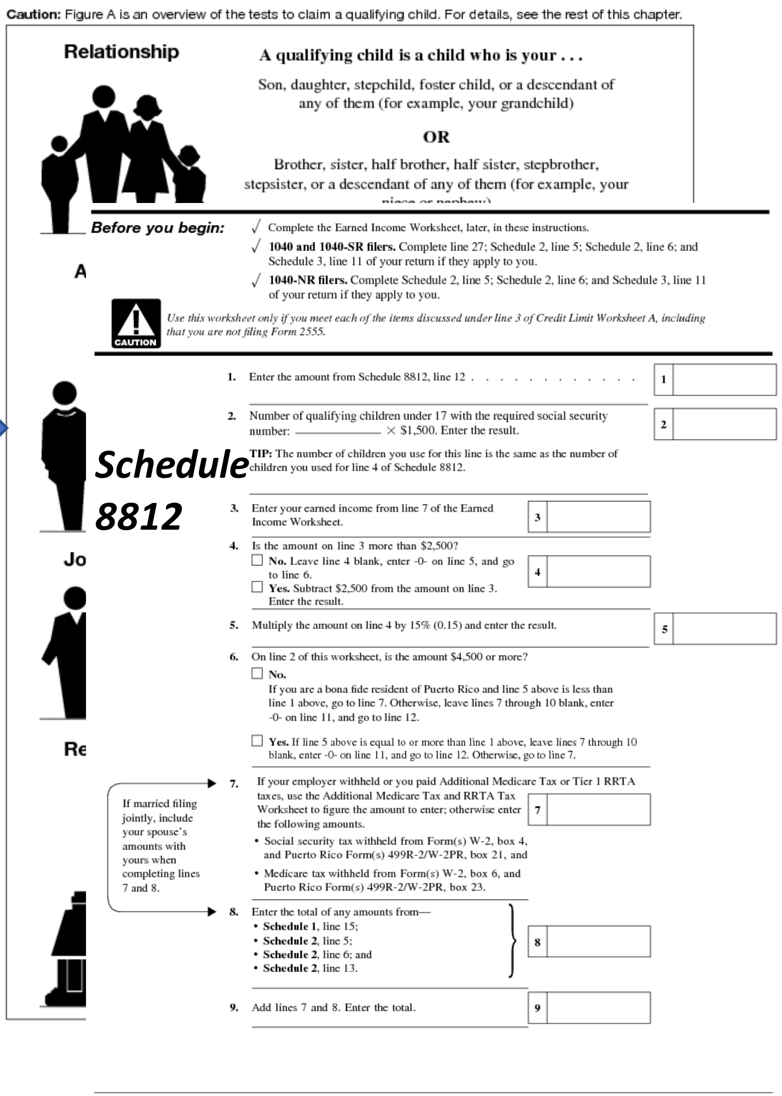
INSTRUCTIONS | Including the instructions for Schedules 1 through 3

2022 Changes

- **Form 1040** has new lines.
- **Schedule 1** has new lines.
- **Filing status name** changed from Qualifying widow(er) to Qualifiying surviving spouse.

For details on these and other changes, see *What's New* in these instructions.

Publication 596 (EITC)



Caution: Figure A is an overview of the tests to claim a qualifying child. For details, see the rest of this chapter.

Relationship

A qualifying child is a child who is your . . .

Son, daughter, stepchild, foster child, or a descendant of any of them (for example, your grandchild)

OR

Brother, sister, half brother, half sister, stepbrother, stepsister, or a descendant of any of them (for example, your niece or nephew)

A

Jo

Re

Schedule 8812

Before you begin:

√ Complete the Earned Income Worksheet, later, in these instructions.

√ **1040 and 1040-SR filers.** Complete line 27; Schedule 2, line 5; Schedule 2, line 6; and Schedule 3, line 11 of your return if they apply to you.

√ **1040-NR filers.** Complete Schedule 2, line 5; Schedule 2, line 6; and Schedule 3, line 11 of your return if they apply to you.

CAUTION *Use this worksheet only if you meet each of the items discussed under line 3 of Credit Limit Worksheet A, including that you are not filing Form 2555.*

1. Enter the amount from Schedule 8812, line 12 . . . . **1** ______

2. Number of qualifying children under 17 with the required social security number: ________ × $1,500. Enter the result. **2** ______

TIP: The number of children you use for this line is the same as the number of children you used for line 4 of Schedule 8812.

3. Enter your earned income from line 7 of the Earned Income Worksheet. **3** ______

4. Is the amount on line 3 more than $2,500?

□ **No.** Leave line 4 blank, enter -0- on line 5, and go to line 6.

□ **Yes.** Subtract $2,500 from the amount on line 3. Enter the result. **4** ______

5. Multiply the amount on line 4 by 15% (0.15) and enter the result. **5** ______

6. On line 2 of this worksheet, is the amount $4,500 or more?

□ **No.**
If you are a bona fide resident of Puerto Rico and line 5 above is less than line 1 above, go to line 7. Otherwise, leave lines 7 through 10 blank, enter -0- on line 11, and go to line 12.

□ **Yes.** If line 5 above is equal to or more than line 1 above, leave lines 7 through 10 blank, enter -0- on line 11, and go to line 12. Otherwise, go to line 7.

If married filing jointly, include your spouse's amounts with yours when completing lines 7 and 8.

7. If your employer withheld or you paid Additional Medicare Tax or Tier 1 RRTA taxes, use the Additional Medicare Tax and RRTA Tax Worksheet to figure the amount to enter; otherwise enter the following amounts. **7** ______

- Social security tax withheld from Form(s) W-2, box 4, and Puerto Rico Form(s) 499R-2/W-2PR, box 21, and
- Medicare tax withheld from Form(s) W-2, box 6, and Puerto Rico Form(s) 499R-2/W-2PR, box 23.

8. Enter the total of any amounts from—
- **Schedule 1**, line 15;
- **Schedule 2**, line 5;
- **Schedule 2**, line 6; and
- **Schedule 2**, line 13.

8 ______

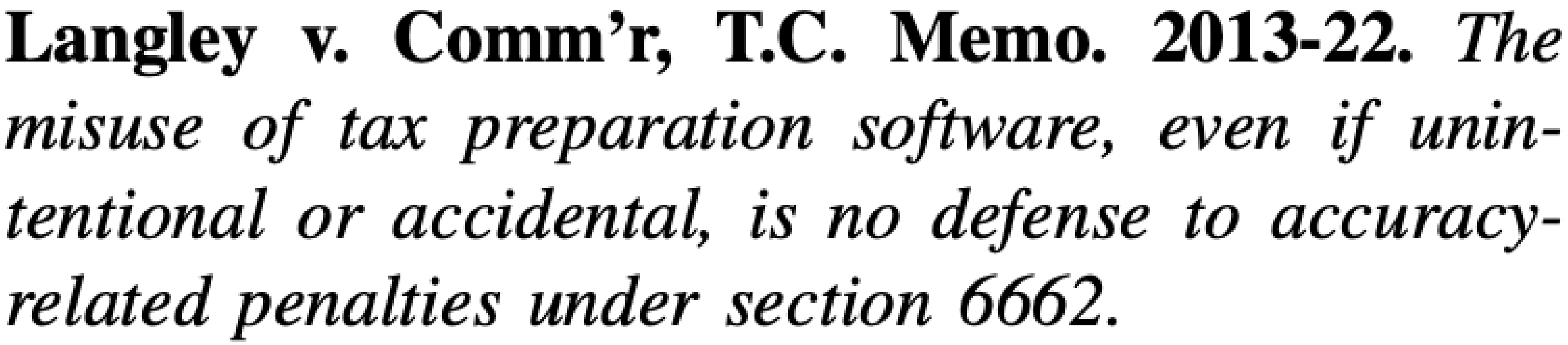
9. Add lines 7 and 8. Enter the total. **9** ______

Tax Preparation Software (US-based)

TaxSlayer®
OpenTaxSolver
Trusted by Millions. Maximum Refund Guaranteed.
Maximize Your Deductions
✓ Easy for first-time users
✓ Mortgage interest, medical and education expenses
Federal & State Returns
plus Federal E-File
Premier
Investments & Rental Property
• Stocks & bonds
• Mutual funds
• Employee stock plans
• Rental income & expenses
• Plus, everything in Deluxe
INCLUDES
✓ Guidance for 1st Time Users
✓ 100% Guaranteed Accurate Calculations
and more!
Works with Quicken
for Windows and Mac

Open-source Tax Solver (OTS)

- 72 million tax returns via software
- 11.2 billion dollars industry
- Free (Open-source) options for low-income



Langley v. Comm’r, T.C. Memo. 2013-22. *The misuse of tax preparation software, even if unintentional or accidental, is no defense to accuracy-related penalties under section 6662.*

Accountable Tax Software

- Comply with laws, regulations, or public policies as they evolve over time.

- Approaches for Accountability of Software
  - Formal verification to ensure compliance;
  - Methodologies for software design, development, and maintenance; and
  - Specification and reasoning about software compliance and accountability.

See solicitation for *Designing Accountable Software Systems* (DASS): https://www.nsf.gov/pubs/2022/nsf22512/nsf22512.htm

Challenges

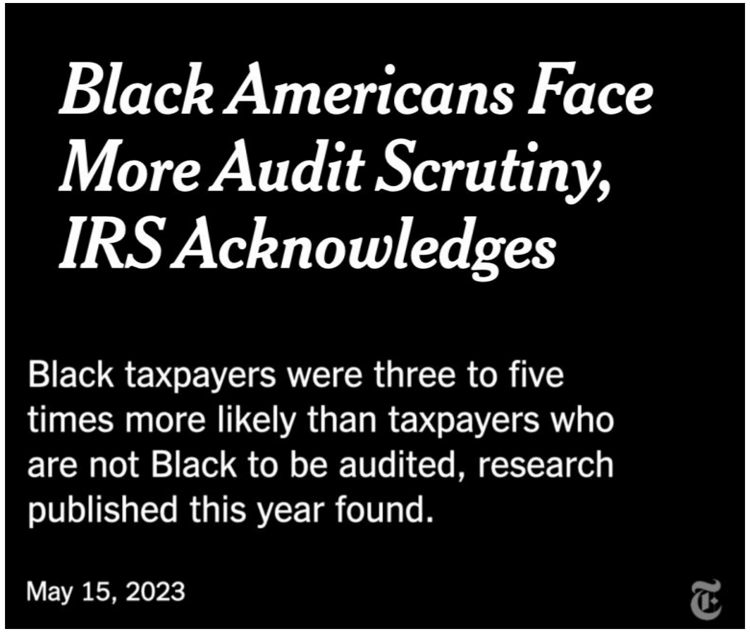
- **Absence of Oracle**
  - Given a taxpayer profile, the ground truth for the tax returns, eligibilities, and credits are not known a prior even for the tax experts;
- **Lack of Trustworthy Dataset**
  - Due to obvious privacy and legal concerns; and
- **Computationally difficult**
  - Finding similar tax profiles is hard (scale, notion of similarity, etc).

Differential Debugging of Tax Software

- **Observation 1**:
  - Tax law adheres to the principles of ``common’’ law;
  - It implements the legal doctrine of precedent; hence,
  - Similar cases must follow similar rulings.

- **Observation 2**:
  - *Horizontal equity* in taxation: relation between similarly situated tax-payers;
  - *Vertical equity* in taxation: relation between taxpayers in different income buckets

Equity in Tax Domain Goes Beyond Software



Black Americans Face More Audit Scrutiny, IRS Acknowledges

Black taxpayers were three to five times more likely than taxpayers who are not Black to be audited, research published this year found.

May 15, 2023

Racial Bias in IRS Tax Audits

Metamorphic Specifications

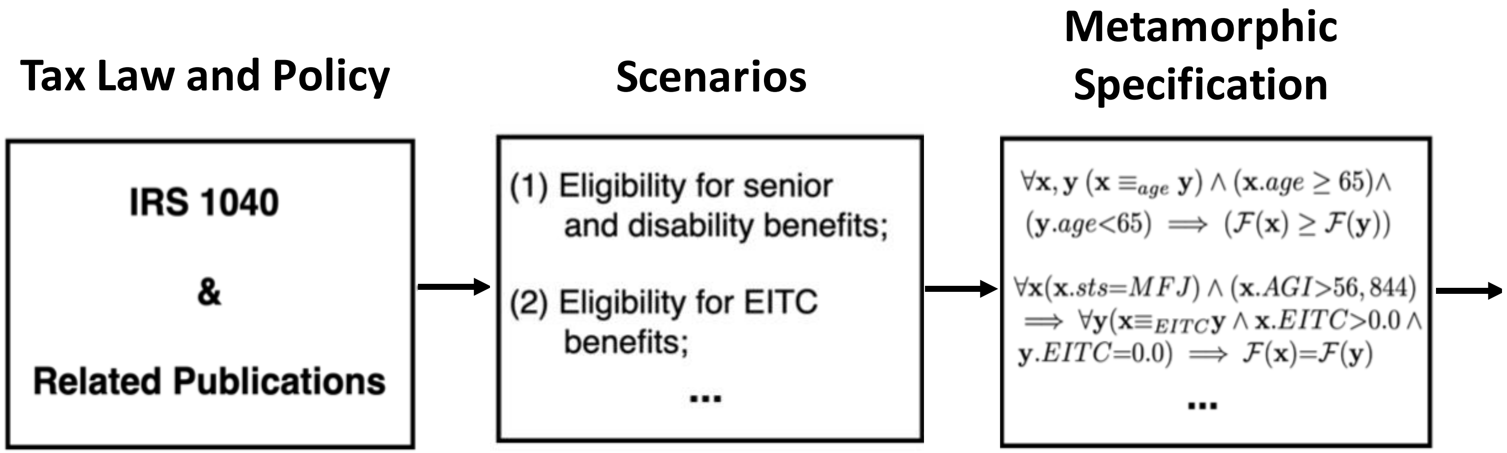
- Validation of software correctness by comparing inputs/outputs

- Example 1: Search Engine
  - $\forall\, q1, q2.\ q1 \subseteq q2 \Rightarrow Items(q1) \geq Items(q2)$

- Example 2: Numerical Software
  - $\forall\, \theta_1, \theta_2.\ \theta_2 = 2 * \pi + \theta_1 \Rightarrow Sin(\theta_1) == Sin(\theta_2)$

- Example 3: Tax Software
  - $\forall\, x_1, x_2.\ x_2 \equiv_{age} x_1 \wedge x1.age \geq x2.age \Rightarrow Return(x_1) \geq Return(x_2)$

TenForty



Tax Law and Policy

IRS 1040

&

Related Publications

Scenarios

(1) Eligibility for senior and disability benefits;

(2) Eligibility for EITC benefits;

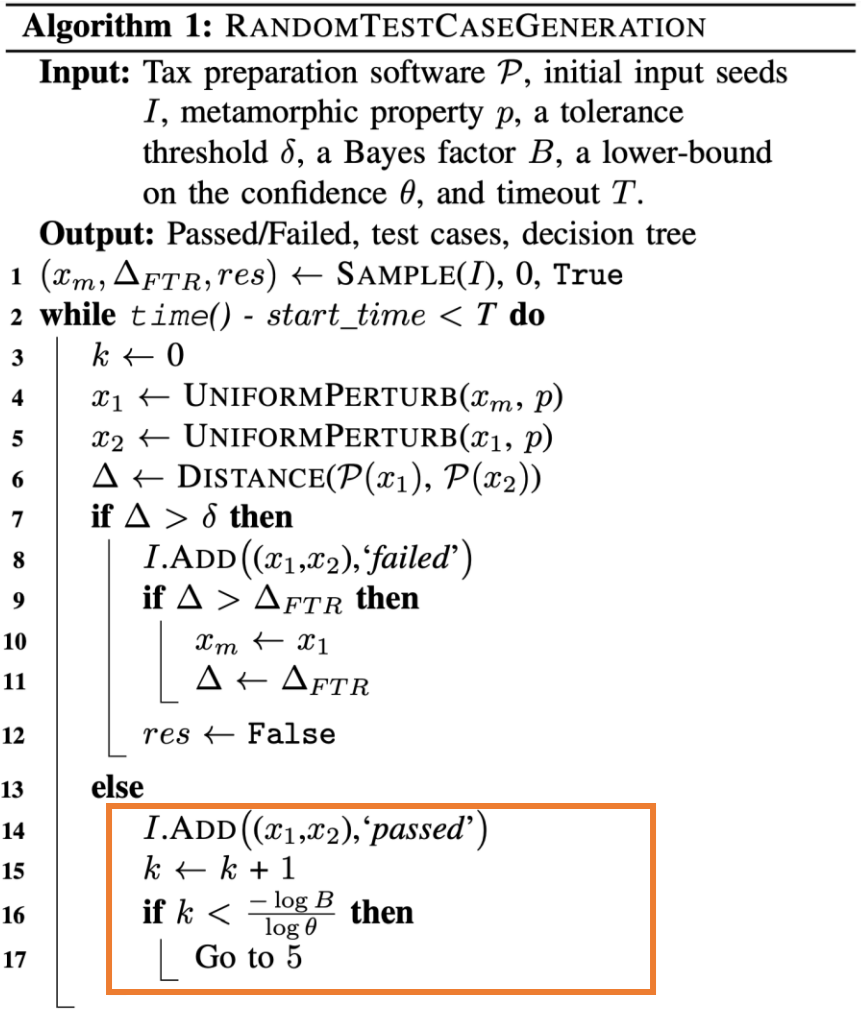
...

Metamorphic Specification

$\forall \mathbf{x},\mathbf{y}\ (\mathbf{x} \equiv_{age} \mathbf{y}) \wedge (\mathbf{x}.age \geq 65) \wedge (\mathbf{y}.age < 65) \implies (\mathcal{F}(\mathbf{x}) \geq \mathcal{F}(\mathbf{y}))$

$\forall \mathbf{x}(\mathbf{x}.sts{=}MFJ) \wedge (\mathbf{x}.AGI{>}56,844) \implies \forall \mathbf{y}(\mathbf{x} \equiv_{EITC} \mathbf{y} \wedge \mathbf{x}.EITC{>}0.0 \wedge \mathbf{y}.EITC{=}0.0) \implies \mathcal{F}(\mathbf{x}){=}\mathcal{F}(\mathbf{y})$

...



Algorithm 1: RandomTestCaseGeneration

Input: Tax preparation software $\mathcal{P}$, initial input seeds I, metamorphic property p, a tolerance threshold δ, a Bayes factor B, a lower-bound on the confidence θ, and timeout T.

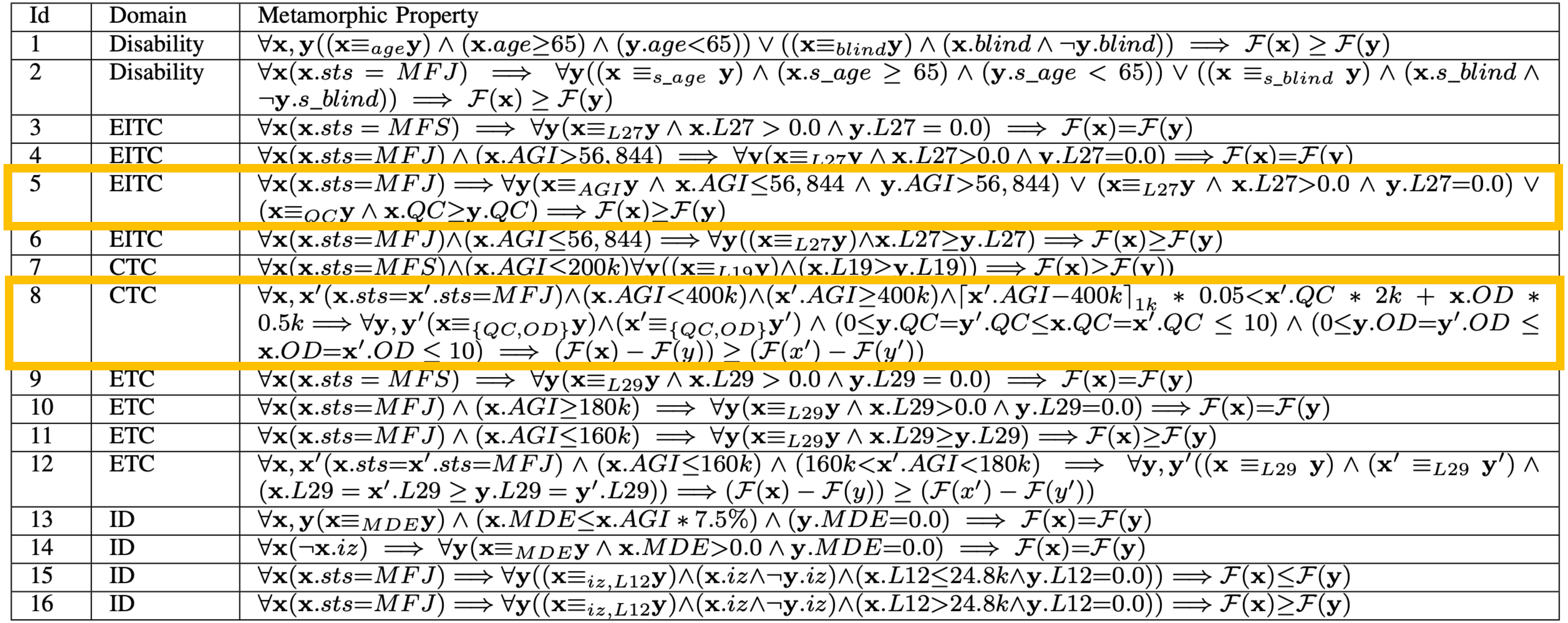
Output: Passed/Failed, test cases, decision tree

```
1  $(x_m, \Delta_{FTR}, res) \leftarrow$ Sample($I$), 0, True
2  while $time()$ - $start\_time < T$ do
3      $k \leftarrow 0$
4      $x_1 \leftarrow$ UniformPerturb($x_m$, $p$)
5      $x_2 \leftarrow$ UniformPerturb($x_1$, $p$)
6      $\Delta \leftarrow$ Distance($\mathcal{P}(x_1)$, $\mathcal{P}(x_2)$)
7      if $\Delta > \delta$ then
8          $I$.Add(($x_1$,$x_2$),'failed')
9          if $\Delta > \Delta_{FTR}$ then
10             $x_m \leftarrow x_1$
11             $\Delta \leftarrow \Delta_{FTR}$
12         $res \leftarrow$ False
13     else
14         $I$.Add(($x_1$,$x_2$),'passed')
15         $k \leftarrow k + 1$
16         if $k < \frac{-\log B}{\log \theta}$ then
17             Go to 5
```

Research Questions

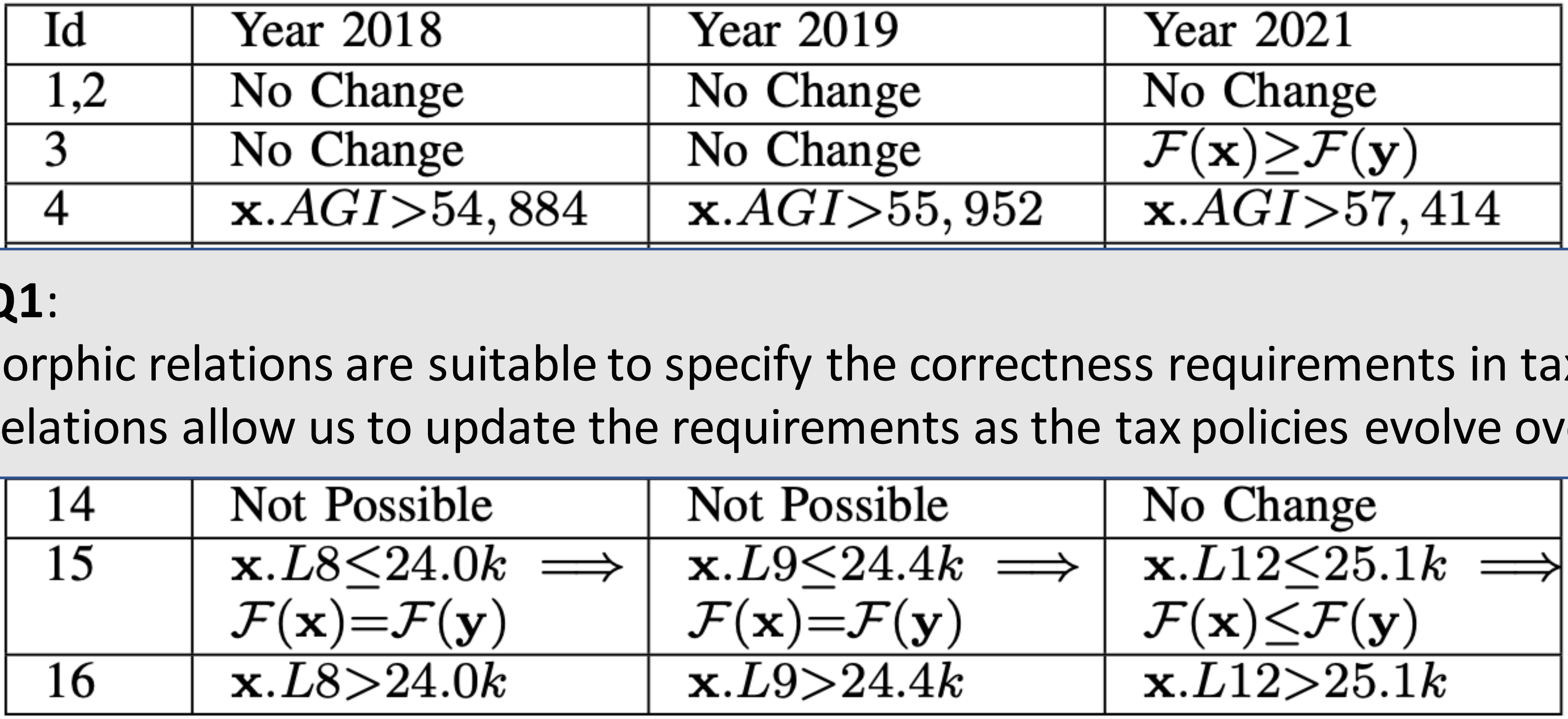
- RQ1: Are **metamorphic relation (MR)** useful to capture the legal requirements of tax preparation software?
- RQ2: Can randomized algorithm with Bayesian guarantees be effective in **testing tax preparation** software against the MR?
- RQ3: Could data-driven fault localization help **pinpoint the root of failures** in the internal and input spaces?

RQ1: Suitability of MR for Tax Law and Policy



Id	Domain	Metamorphic Property
1	Disability	$\forall \mathbf{x}, \mathbf{y}((\mathbf{x} \equiv_{age} \mathbf{y}) \land (\mathbf{x}.age \geq 65) \land (\mathbf{y}.age < 65)) \lor ((\mathbf{x} \equiv_{blind} \mathbf{y}) \land (\mathbf{x}.blind \land \neg \mathbf{y}.blind)) \implies \mathcal{F}(\mathbf{x}) \geq \mathcal{F}(\mathbf{y})$
2	Disability	$\forall \mathbf{x}(\mathbf{x}.sts = MFJ) \implies \forall \mathbf{y}((\mathbf{x} \equiv_{s\_age} \mathbf{y}) \land (\mathbf{x}.s\_age \geq 65) \land (\mathbf{y}.s\_age < 65)) \lor ((\mathbf{x} \equiv_{s\_blind} \mathbf{y}) \land (\mathbf{x}.s\_blind \land \neg \mathbf{y}.s\_blind)) \implies \mathcal{F}(\mathbf{x}) \geq \mathcal{F}(\mathbf{y})$
3	EITC	$\forall \mathbf{x}(\mathbf{x}.sts = MFS) \implies \forall \mathbf{y}(\mathbf{x} \equiv_{L27} \mathbf{y} \land \mathbf{x}.L27 > 0.0 \land \mathbf{y}.L27 = 0.0) \implies \mathcal{F}(\mathbf{x}) = \mathcal{F}(\mathbf{y})$
4	EITC	$\forall \mathbf{x}(\mathbf{x}.sts = MFJ) \land (\mathbf{x}.AGI > 56,844) \implies \forall \mathbf{y}(\mathbf{x} \equiv_{L27} \mathbf{y} \land \mathbf{x}.L27 > 0.0 \land \mathbf{y}.L27 = 0.0) \implies \mathcal{F}(\mathbf{x}) = \mathcal{F}(\mathbf{y})$
5	EITC	$\forall \mathbf{x}(\mathbf{x}.sts = MFJ) \implies \forall \mathbf{y}(\mathbf{x} \equiv_{AGI} \mathbf{y} \land \mathbf{x}.AGI \leq 56,844 \land \mathbf{y}.AGI > 56,844) \lor (\mathbf{x} \equiv_{L27} \mathbf{y} \land \mathbf{x}.L27 > 0.0 \land \mathbf{y}.L27 = 0.0) \lor (\mathbf{x} \equiv_{QC} \mathbf{y} \land \mathbf{x}.QC \geq \mathbf{y}.QC) \implies \mathcal{F}(\mathbf{x}) \geq \mathcal{F}(\mathbf{y})$
6	EITC	$\forall \mathbf{x}(\mathbf{x}.sts = MFJ) \land (\mathbf{x}.AGI \leq 56,844) \implies \forall \mathbf{y}((\mathbf{x} \equiv_{L27} \mathbf{y}) \land \mathbf{x}.L27 \geq \mathbf{y}.L27) \implies \mathcal{F}(\mathbf{x}) \geq \mathcal{F}(\mathbf{y})$
7	CTC	$\forall \mathbf{x}(\mathbf{x}.sts = MFS) \land (\mathbf{x}.AGI < 200k) \forall \mathbf{y}((\mathbf{x} \equiv_{L19} \mathbf{y}) \land (\mathbf{x}.L19 > \mathbf{y}.L19)) \implies \mathcal{F}(\mathbf{x}) > \mathcal{F}(\mathbf{y}))$
8	CTC	$\forall \mathbf{x}, \mathbf{x}'(\mathbf{x}.sts = \mathbf{x}'.sts = MFJ) \land (\mathbf{x}.AGI < 400k) \land (\mathbf{x}'.AGI \geq 400k) \land \lceil \mathbf{x}'.AGI - 400k \rceil_{1k} * 0.05 < \mathbf{x}'.QC * 2k + \mathbf{x}.OD * 0.5k \implies \forall \mathbf{y}, \mathbf{y}'(\mathbf{x} \equiv_{\{QC,OD\}} \mathbf{y}) \land (\mathbf{x}' \equiv_{\{QC,OD\}} \mathbf{y}') \land (0 \leq \mathbf{y}.QC = \mathbf{y}'.QC \leq \mathbf{x}.QC = \mathbf{x}'.QC \leq 10) \land (0 \leq \mathbf{y}.OD = \mathbf{y}'.OD \leq \mathbf{x}.OD = \mathbf{x}'.OD \leq 10) \implies (\mathcal{F}(\mathbf{x}) - \mathcal{F}(y)) \geq (\mathcal{F}(x') - \mathcal{F}(y'))$
9	ETC	$\forall \mathbf{x}(\mathbf{x}.sts = MFS) \implies \forall \mathbf{y}(\mathbf{x} \equiv_{L29} \mathbf{y} \land \mathbf{x}.L29 > 0.0 \land \mathbf{y}.L29 = 0.0) \implies \mathcal{F}(\mathbf{x}) = \mathcal{F}(\mathbf{y})$
10	ETC	$\forall \mathbf{x}(\mathbf{x}.sts = MFJ) \land (\mathbf{x}.AGI \geq 180k) \implies \forall \mathbf{y}(\mathbf{x} \equiv_{L29} \mathbf{y} \land \mathbf{x}.L29 > 0.0 \land \mathbf{y}.L29 = 0.0) \implies \mathcal{F}(\mathbf{x}) = \mathcal{F}(\mathbf{y})$
11	ETC	$\forall \mathbf{x}(\mathbf{x}.sts = MFJ) \land (\mathbf{x}.AGI \leq 160k) \implies \forall \mathbf{y}(\mathbf{x} \equiv_{L29} \mathbf{y} \land \mathbf{x}.L29 \geq \mathbf{y}.L29) \implies \mathcal{F}(\mathbf{x}) \geq \mathcal{F}(\mathbf{y})$
12	ETC	$\forall \mathbf{x}, \mathbf{x}'(\mathbf{x}.sts = \mathbf{x}'.sts = MFJ) \land (\mathbf{x}.AGI \leq 160k) \land (160k < \mathbf{x}'.AGI < 180k) \implies \forall \mathbf{y}, \mathbf{y}'((\mathbf{x} \equiv_{L29} \mathbf{y}) \land (\mathbf{x}' \equiv_{L29} \mathbf{y}') \land (\mathbf{x}.L29 = \mathbf{x}'.L29 \geq \mathbf{y}.L29 = \mathbf{y}'.L29)) \implies (\mathcal{F}(\mathbf{x}) - \mathcal{F}(y)) \geq (\mathcal{F}(x') - \mathcal{F}(y'))$
13	ID	$\forall \mathbf{x}, \mathbf{y}(\mathbf{x} \equiv_{MDE} \mathbf{y}) \land (\mathbf{x}.MDE \leq \mathbf{x}.AGI * 7.5\%) \land (\mathbf{y}.MDE = 0.0) \implies \mathcal{F}(\mathbf{x}) = \mathcal{F}(\mathbf{y})$
14	ID	$\forall \mathbf{x}(\neg \mathbf{x}.iz) \implies \forall \mathbf{y}(\mathbf{x} \equiv_{MDE} \mathbf{y} \land \mathbf{x}.MDE > 0.0 \land \mathbf{y}.MDE = 0.0) \implies \mathcal{F}(\mathbf{x}) = \mathcal{F}(\mathbf{y})$
15	ID	$\forall \mathbf{x}(\mathbf{x}.sts = MFJ) \implies \forall \mathbf{y}((\mathbf{x} \equiv_{iz,L12} \mathbf{y}) \land (\mathbf{x}.iz \land \neg \mathbf{y}.iz) \land (\mathbf{x}.L12 \leq 24.8k \land \mathbf{y}.L12 = 0.0)) \implies \mathcal{F}(\mathbf{x}) \leq \mathcal{F}(\mathbf{y})$
16	ID	$\forall \mathbf{x}(\mathbf{x}.sts = MFJ) \implies \forall \mathbf{y}((\mathbf{x} \equiv_{iz,L12} \mathbf{y}) \land (\mathbf{x}.iz \land \neg \mathbf{y}.iz) \land (\mathbf{x}.L12 > 24.8k \land \mathbf{y}.L12 = 0.0)) \implies \mathcal{F}(\mathbf{x}) \geq \mathcal{F}(\mathbf{y})$

RQ1: Suitability of MR for Tax Law and Policy

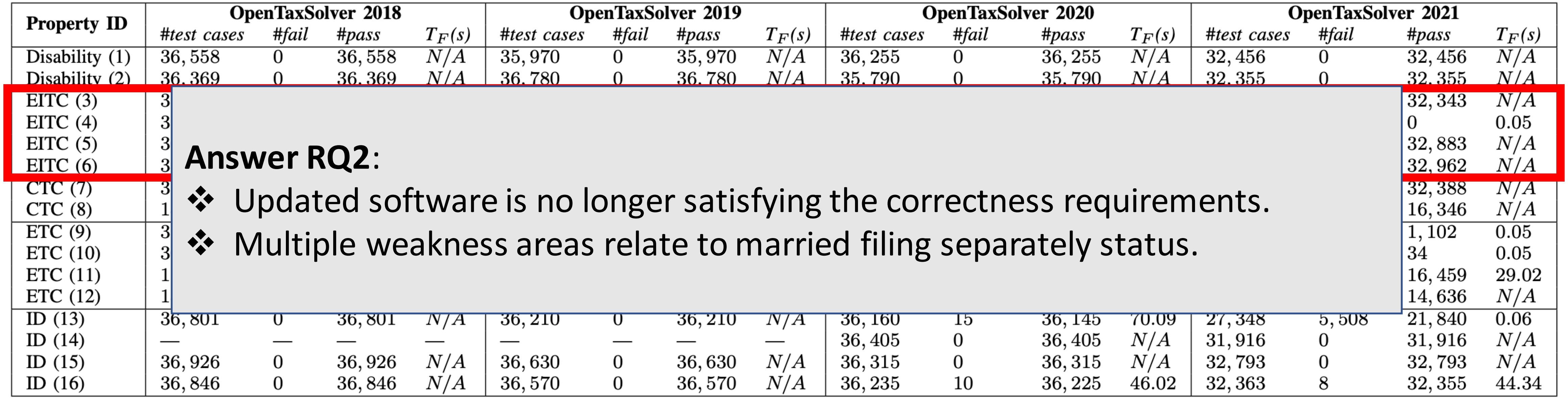


Id	Year 2018	Year 2019	Year 2021
1,2	No Change	No Change	No Change
3	No Change	No Change	$\mathcal{F}(\mathbf{x})\geq\mathcal{F}(\mathbf{y})$
4	$\mathbf{x}.AGI>54,884$	$\mathbf{x}.AGI>55,952$	$\mathbf{x}.AGI>57,414$
14	Not Possible	Not Possible	No Change
15	$\mathbf{x}.L8\leq24.0k \implies \mathcal{F}(\mathbf{x})=\mathcal{F}(\mathbf{y})$	$\mathbf{x}.L9\leq24.4k \implies \mathcal{F}(\mathbf{x})=\mathcal{F}(\mathbf{y})$	$\mathbf{x}.L12\leq25.1k \implies \mathcal{F}(\mathbf{x})\leq\mathcal{F}(\mathbf{y})$
16	$\mathbf{x}.L8>24.0k$	$\mathbf{x}.L9>24.4k$	$\mathbf{x}.L12>25.1k$

Answer RQ1:

- ❖ Metamorphic relations are suitable to specify the correctness requirements in tax software.
- ❖ These relations allow us to update the requirements as the tax policies evolve over time.

RQ2: Testing Software against MR requirements

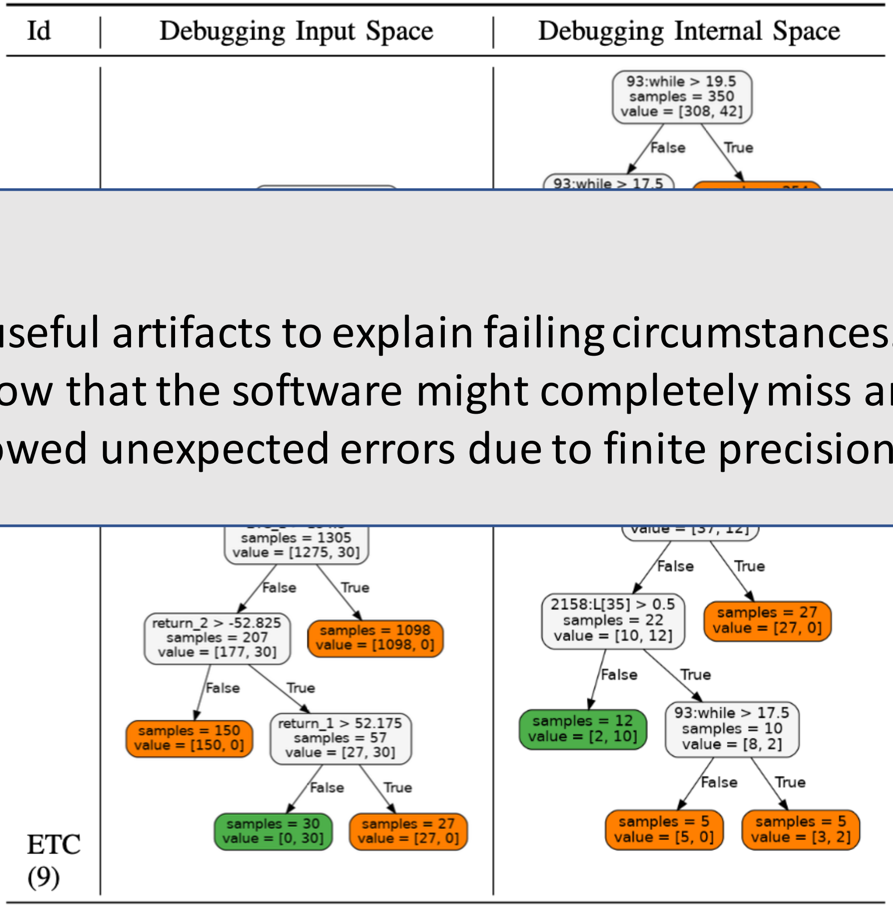


Property ID	OpenTaxSolver 2018				OpenTaxSolver 2019				OpenTaxSolver 2020				OpenTaxSolver 2021			
	#test cases	*#fail*	*#pass*	$T_F(s)$	*#test cases*	*#fail*	*#pass*	$T_F(s)$	*#test cases*	*#fail*	*#pass*	$T_F(s)$	*#test cases*	*#fail*	*#pass*	$T_F(s)$
Disability (1)	36,558	0	36,558	N/A	35,970	0	35,970	N/A	36,255	0	36,255	N/A	32,456	0	32,456	N/A
Disability (2)	36,369	0	36,369	N/A	36,780	0	36,780	N/A	35,790	0	35,790	N/A	32,355	0	32,355	N/A
EITC (3)	3														32,343	N/A
EITC (4)	3														0	0.05
EITC (5)	3														32,883	N/A
EITC (6)	3														32,962	N/A
CTC (7)	3														32,388	N/A
CTC (8)	1														16,346	N/A
ETC (9)	3														1,102	0.05
ETC (10)	3														34	0.05
ETC (11)	1														16,459	29.02
ETC (12)	1														14,636	N/A
ID (13)	36,801	0	36,801	N/A	36,210	0	36,210	N/A	36,160	15	36,145	70.09	27,348	5,508	21,840	0.06
ID (14)	—	—	—	—	—	—	—	—	36,405	0	36,405	N/A	31,916	0	31,916	N/A
ID (15)	36,926	0	36,926	N/A	36,630	0	36,630	N/A	36,315	0	36,315	N/A	32,793	0	32,793	N/A
ID (16)	36,846	0	36,846	N/A	36,570	0	36,570	N/A	36,235	10	36,225	46.02	32,363	8	32,355	44.34

Answer RQ2:

- ❖ Updated software is no longer satisfying the correctness requirements.
- ❖ Multiple weakness areas relate to married filing separately status.

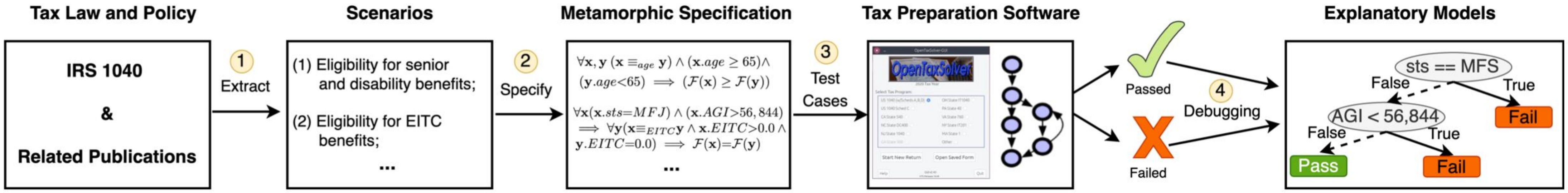
RQ3: Data-Driven Root Cause Identification



Answer RQ3:

- ❖ Decision trees are useful artifacts to explain failing circumstances.
- ❖ Our experiences show that the software might completely miss an eligibility condition.
- ❖ Our results also showed unexpected errors due to finite precision in the computation.

Metamorphic Testing and Debugging of Tax Preparation Software (https://arxiv.org/abs/2205.04998)


Tax Law and Policy
IRS 1040
&
Related Publications
1 Extract
Scenarios
(1) Eligibility for senior and disability benefits;
(2) Eligibility for EITC benefits;
...
2 Specify
Metamorphic Specification
$\forall \mathbf{x}, \mathbf{y}\ (\mathbf{x} \equiv_{age} \mathbf{y}) \land (\mathbf{x}.age \geq 65) \land (\mathbf{y}.age < 65) \implies (\mathcal{F}(\mathbf{x}) \geq \mathcal{F}(\mathbf{y}))$
$\forall \mathbf{x} (\mathbf{x}.sts = MFJ) \land (\mathbf{x}.AGI > 56,844) \implies \forall \mathbf{y} (\mathbf{x} \equiv_{EITC} \mathbf{y} \land \mathbf{x}.EITC > 0.0 \land \mathbf{y}.EITC = 0.0) \implies \mathcal{F}(\mathbf{x}) = \mathcal{F}(\mathbf{y})$
...
3 Test Cases
Tax Preparation Software
OpenTaxSolver
Passed
Failed
4 Debugging
Explanatory Models
sts == MFS
False
True
Fail
AGI < 56,844
False
True
Pass
Fail

Forensic DNA Software

- New York City's Office of Chief Medical Examiner (OCME) for thousands of criminal cases between 2011 and 2017
- Undisclosed data dropping method CheckFrequencyForRemoval()
- Falsely skew results toward false inclusion for individuals whose DNA was not present.

"Do I Qualify?" Screening Software

- Poverty management systems in Pennsylvania (Check Eligibility)
- Comparative implementation of benefit eligibility handbook
- Errors in the eligibility checking: Exclude the most vulnerable families from receiving the essential aids